Unit 5

Linear algebra

Introduction

Figure 1 shows the network of water pipes connecting the pump and five taps that control the heating system in a (fictional) greenhouse. The engineer who maintains the pump can determine the pressure P at which hot water is supplied to the system, and the tightness T_i ($i = 1, \ldots, 5$) of each tap. He knows from experience that making the system perform as required throughout the day is no easy matter. The effect of each part of the system depends on the rate at which hot water flows through it. Because the system is closed, all the water that flows into any junction must flow out again. As a result, there are only three key rates of flow, namely f_1, f_2 and f_3, that together determine the rate of flow in every part of the system (see Figure 1). However, the equations that relate these flow rates to the supply pressure and tap settings are not simple.

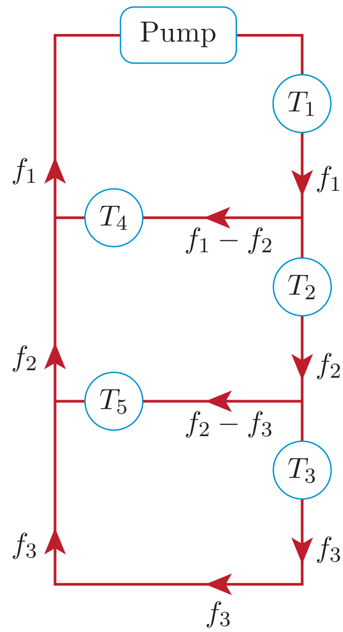


Figure 1 The pump and network of taps that control a hot water heating system

The engineer knows that the system is described by the following simultaneous equations (which you are not expected to derive):

$$\begin{aligned}
&T_1 f_1 + T_4 f_1 - T_4 f_2 = P,\\
&T_2 f_2 + T_5 f_2 - T_5 f_3 - T_4 f_1 + T_4 f_2 = 0,\\
&T_3 f_3 - T_5 f_2 + T_5 f_3 = 0.
\end{aligned}$$

These equations can be expressed in terms of matrices as

$$\begin{bmatrix} T_1 + T_4 & -T_4 & 0 \\ -T_4 & T_2 + T_4 + T_5 & -T_5 \\ 0 & -T_5 & T_3 + T_5 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} P \\ 0 \\ 0 \end{bmatrix}.$$

Solving this matrix equation is typical of the kind of problem that is best treated using the methods of linear algebra.

Linear algebra is a branch of mathematics that grew out of the solution of systems of equations like that given above. The equations are said to be *linear* because the variables of interest, in this case the flow rates f_1, f_2 and f_3, appear only to the first degree in each term; there are no powers of flow rates, such as f_1^2 or f_2^3, nor are there any products of flow rates, such as $f_1 f_2$ or $f_2 f_3$. Although there are several variables, the equations are a generalisation of the kind of *linear* equation that describes a straight line in two-dimensional Cartesian coordinates, $y = mx + c$. This 'linearity' makes the equations well suited to representation by matrices, and once they are in matrix form, solving them is assisted by everything that mathematicians have learned from their studies of matrix algebra and determinants. Linear algebra is the subject that gathers together those insights and methods, and generalises them to create a coherent body of mathematical knowledge rather than an assortment of calculational tricks.

Linear algebra is an important topic in modern mathematics. It can be approached in many different ways, ranging from the very abstract, in which there is little or no mention of linear equations, matrices or determinants, to the very practical, in which there is a concentration on matrix methods and the practicalities of computer-based calculations.

In this unit we aim to steer a middle course, emphasising the language of matrices and determinants while providing an introduction to those parts of the subject that are most relevant to applications without getting bogged down in calculational details. With this aim in mind, we will concentrate on two important approaches to problem solving that are characteristic of linear algebra. The first is the use of 'row operations', which form the basis of a range of methods for solving systems of linear equations. The second is the subject of 'eigenvalues' and 'eigenvectors', which are numbers and vectors associated with a matrix. As you will see, knowing the eigenvalues and eigenvectors of a matrix can greatly assist the solution of problems involving that matrix.

The prefix *eigen* comes from German and indicates 'inherent', as these things are inherent to a matrix.

Study guide

Section 1 shows how to use row operations to solve a system of n simultaneous equations of the form $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is a given $n \times n$ matrix, $\mathbf{b}$ is a given column vector with n elements, and we are required to find $\mathbf{x}$, the n-element column vector that represents the solution. The method that we will use is called *Gaussian elimination*.

Section 2 addresses the question of determining the result $\mathbf{x}_k = \mathbf{A}^k\mathbf{x}_0$ of k applications of a square matrix $\mathbf{A}$ to an initial column vector $\mathbf{x}_0$. This leads us to study the equation $\mathbf{Ax} = \lambda\mathbf{x}$, where the scalar λ is called an *eigenvalue* of the matrix $\mathbf{A}$, with $\mathbf{x}$ being the corresponding *eigenvector*.

Section 3 gives the general method for finding eigenvalues and eigenvectors. (This will be of particular use in the next unit.)

In each of Sections 1 and 3, there are procedures for solving problems in linear algebra. Take care that you understand the examples given for these procedures and can do the corresponding exercises; these give a good idea of the assessable learning outcomes of this unit.

Section 4, which is optional, contains some further results concerning the eigenvalues and eigenvectors of *symmetric* matrices.

1 Linear algebra, row operations and Gaussian elimination

1.1 Linear equations and their manipulation

You should already be familiar with the notion of a linear equation as a generalisation of $y = mx + c$, but let us start with a definition so that there can be no doubt.

Linear equation

An equation involving n variables $x_1, x_2, \ldots, x_n$ is said to be **linear** in each of those variables if it can be written in the form

$$a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n = d,$$

where $a_1, a_2, \ldots, a_n$ and d are constants.

Example 1

Show that the Cartesian equation of a straight line, $y = mx + c$, is a linear equation by identifying each of the variables and constants.

Solution

The equation can be rewritten in the form $y - mx = c$. If we compare this with the general form of a linear equation of $n = 2$ variables, we can see that the rewritten Cartesian equation is linear if we make the identifications $a_1 = 1$, $x_1 = y$, $a_2 = -m$, $x_2 = x$ and $d = c$.

Exercise 1

The Cartesian equation of a plane that passes through the origin of a three-dimensional system of coordinates is $ax + by + cz = 0$, where a, b and c are constants. Is this equation linear in the variables x, y and z? Justify your answer.

A very simple system of linear equations to solve would consist of the two equations

$$3x + 2y = 8,$$
$$x - y = 1.$$

In order to solve them, we should try to isolate one of the variables, x or y. In this case that can easily be done by adding twice the second equation to the first. This eliminates y and produces the equation

$$5x = 10,$$

implying that $x = 2$. Substituting this back into either of the original equations immediately shows that $y = 1$.

What follows will not always be so easy, but much of this first section will essentially be a systematic elaboration of the manipulation that has just been carried out.

Consider first a system of three simultaneous linear equations, (E_1), (E_2) and (E_3):

$$\begin{aligned} x_1 - 4x_2 + 2x_3 &= -9, &\qquad (E_1)\\ 3x_1 - 2x_2 + 3x_3 &= 7, &\qquad (E_2)\\ 8x_1 - 2x_2 + 9x_3 &= 34, &\qquad (E_3) \end{aligned}$$

with three unknowns, x_1, x_2 and x_3.

We may obtain a system of two simultaneous equations, for x_2 and x_3, by subtracting suitable multiples of (E_1) from (E_2) and (E_3). In this example, we may subtract 3 times (E_1) from (E_2), to obtain

$$10x_2 - 3x_3 = 34. \qquad (E_{2a}) = (E_2) - 3(E_1)$$

(Note that we have shown on the right the manipulations required to obtain the equation on the left.) We may also subtract 8 times (E_1) from (E_3), to obtain

$$30x_2 - 7x_3 = 106. \qquad (E_{3a}) = (E_3) - 8(E_1)$$

So now we have two simultaneous equations, neither of which involves x_1.

Next we may subtract 3 times (E_{2a}) from (E_{3a}), to obtain

$$2x_3 = 4, \qquad (E_{3b}) = (E_{3a}) - 3(E_{2a})$$

which tells us that $x_3 = 4/2 = 2$. Substituting this result back into (E_{2a}) tells us that $x_2 = (3x_3 + 34)/10 = 4$, and substituting the values for x_3 and x_2 back into (E_1) tells us that $x_1 = 4x_2 - 2x_3 - 9 = 3$. So the complete solution of the system is $x_1 = 3$, $x_2 = 4$, $x_3 = 2$.

Next we examine a convenient way to represent this process, using matrix notation.

1.2 The augmented matrix and row operations

Given a set of linear equations, the first thing to do is write them in a standard way, with the constants on the right and the three variable terms on the left. Equations (E_1), (E_2) and (E_3) above are written in this way. Notice how the variable terms in x_1, x_2 and x_3 are lined up in separate vertical columns. If there are any missing terms (where the coefficient is zero), leave a gap in the column. The form of equations (E_1), (E_2) and (E_3) immediately suggests that they can be expressed as a single matrix equation

$$\begin{bmatrix} 1 & -4 & 2\\ 3 & -2 & 3\\ 8 & -2 & 9 \end{bmatrix} \begin{bmatrix} x_1\\ x_2\\ x_3 \end{bmatrix} = \begin{bmatrix} -9\\ 7\\ 34 \end{bmatrix}. \qquad (1)$$

The correctness of this suggestion can be confirmed by working out the matrix multiplication on the left-hand side. This gives

$$\begin{bmatrix} x_1 - 4x_2 + 2x_3\\ 3x_1 - 2x_2 + 3x_3\\ 8x_1 - 2x_2 + 9x_3 \end{bmatrix} = \begin{bmatrix} -9\\ 7\\ 34 \end{bmatrix}.$$

The matrix on the left can equal the matrix on the right only if corresponding elements are equal, which gives the three equations (E_1), (E_2) and (E_3).

We can represent the matrix equation (1) symbolically by

$$\mathbf{Ax} = \mathbf{b}, \quad \text{where } \mathbf{A} = \begin{bmatrix} 1 & -4 & 2 \\ 3 & -2 & 3 \\ 8 & -2 & 9 \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} -9 \\ 7 \\ 34 \end{bmatrix}.$$

Remember that the emboldening of printed symbols, such as $\mathbf{A}$, $\mathbf{x}$, $\mathbf{b}$, can be indicated in handwritten work by using an underline, as in $\underline{A}$, $\underline{x}$, $\underline{b}$. It is particularly important to remember this when using a symbol to represent a vector quantity.

Any system of linear equations may be represented in a similar way, even if some initial rearrangement is needed to achieve the required layout. Once the equations have been expressed in matrix form it is conventional to call $\mathbf{A}$ the **coefficient matrix**, $\mathbf{x}$ the **unknown vector**, and $\mathbf{b}$ the **right-hand-side vector** or sometimes the **constant vector** since it represents the terms that are independent of the variables.

Exercise 2

Write each of the following systems in the matrix form $\mathbf{Ax} = \mathbf{b}$.

(a) $$\begin{aligned} x_1 + x_2 - x_3 &= 2, \\ 5x_1 + 2x_2 + 2x_3 &= 20, \\ 4x_1 - 2x_2 - 3x_3 &= 15. \end{aligned}$$

(b) $$\begin{aligned} -2 + x_2 - x_3 &= -x_1, \\ -2x_3 - 2x_2 &= 5x_1 - 20, \\ -2x_2 - 3x_3 &= 15 - 4x_1. \end{aligned}$$

(c) $$\begin{aligned} 2x + 3y - 4z &= 0, \\ 2x \qquad + 3z &= 3, \\ 6y - 2z &= 0. \end{aligned}$$

Using matrix notation, we could represent each of the intermediate stages of the calculation in Subsection 1.1 as a matrix equation. All we would have to do is to change the coefficient matrix and the right-hand-side vector in an appropriate way at each stage. You are welcome to check this for yourself, but doing so is really something of a waste of time since it would involve a lot of repetitive writing that would not achieve anything new. A better approach is to introduce a new 'shorthand' that captures just the essential information and the steps in the calculation. To do this, we introduce a notational device called the **augmented matrix** that combines the elements of the coefficient matrix and the right-hand-side vector. For the matrix equation (1) given above, the augmented matrix is written as

$$\mathbf{A}|\mathbf{b} = \left[\begin{array}{ccc|c} 1 & -4 & 2 & -9 \\ 3 & -2 & 3 & 7 \\ 8 & -2 & 9 & 34 \end{array}\right]$$

Don't look for any deep significance in the augmented matrix. It's not a new 'kind' of matrix, just a useful means of efficiently displaying the crucial information in a system of linear equations. To show how that

information changes as the equations are manipulated, it is helpful to label each row. We do this with a bold $\mathbf{R}$ (for row) and a subscript as follows:

$$\mathbf{A}|\mathbf{b} = \left[\begin{array}{ccc|c} 1 & -4 & 2 & -9 \\ 3 & -2 & 3 & 7 \\ 8 & -2 & 9 & 34 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array}.$$

Using these notational devices, we can compactly indicate any step in the manipulation of the simultaneous equations. For instance, suppose that we want to change the second equation by subtracting three times the first equation from it. We can show this using the augmented matrix by doing the following steps.

1. Write down row $\mathbf{R}_1$ unchanged.
2. Indicate the planned change as $\mathbf{R}_2 - 3\mathbf{R}_1$, and write that to the left of the second row.
3. Write the values resulting from the change in the second row.
4. Relabel the second row as $\mathbf{R}_{2a}$.
5. Write down row $\mathbf{R}_3$ unchanged.

So in this particular case we get

$$\begin{array}{r} \\ \mathbf{R}_2 - 3\mathbf{R}_1 \\ \\ \end{array} \left[\begin{array}{ccc|c} 1 & -4 & 2 & -9 \\ 0 & 10 & -3 & 34 \\ 8 & -2 & 9 & 34 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_3 \end{array}.$$

Having just reduced the coefficient of x_1 to 0 in the second row, the next step in solving the simultaneous equations is to make a similar reduction in the third row. This is again achieved by subtracting (or adding) an appropriate multiple of another entire row. In this case we subtract 8 times $\mathbf{R}_1$ from $\mathbf{R}_3$, indicating this as follows:

$$\begin{array}{r} \\ \\ \mathbf{R}_3 - 8\mathbf{R}_1 \end{array} \left[\begin{array}{ccc|c} 1 & -4 & 2 & -9 \\ 0 & 10 & -3 & 34 \\ 0 & 30 & -7 & 106 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array}.$$

Here we are again showing the steps that must be taken on the left, the resulting augmented matrix in the middle, and the updated row labels on the right. Our next step is to reduce to zero the coefficient of x_2 in the third row without upsetting the pattern of zeros in the first column. This is achieved as follows:

Note that we had to subtract a multiple of $\mathbf{R}_{2a}$ here. Had we tried to subtract a multiple of $\mathbf{R}_1$ from $\mathbf{R}_{3a}$, we would have gained an unwanted non-zero entry in the first column of the third row.

$$\begin{array}{r} \\ \\ \mathbf{R}_{3a} - 3\mathbf{R}_{2a} \end{array} \left[\begin{array}{ccc|c} 1 & -4 & 2 & -9 \\ 0 & 10 & -3 & 34 \\ 0 & 0 & 2 & 4 \end{array}\right].$$

Since we have completed all the row operations involved in our particular problem, we have deliberately omitted an updated set of row labels to the right of the final augmented matrix, though we could easily have included them if they were needed.

Note that all the operations that we have recorded on the left have involved subtracting multiples of one row from another row. These are all examples of what are generally referred to as **row operations** on the augmented matrix. The full range of allowable row operations mimics

those steps that we might have taken in solving the original system of equations. They can all be built from successive combinations of the following three *elementary row operations*.

Elementary row operations

- Interchange any two complete rows ($\mathbf{R}_i \leftrightarrow \mathbf{R}_j$).
- Multiply each element in a row by the same constant ($\mathbf{R}_i \to \lambda\mathbf{R}_i$).
- Add (or subtract) a constant multiple of one row to (or from) another row ($\mathbf{R}_i \to \mathbf{R}_i \pm \lambda\mathbf{R}_j$).

Note that row operations always involve *entire* rows of the augmented matrix, thus including the entries from the right-hand-side vector as well as those from the coefficient matrix.

The overall effect of the row operations involved in our example has been to produce a triangle of zeros in the bottom left-hand corner of the augmented matrix. If we use that augmented matrix to write down the final set of linear equations in matrix form, we obtain

$$\begin{bmatrix} 1 & -4 & 2 \\ 0 & 10 & -3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -9 \\ 34 \\ 4 \end{bmatrix}.$$

The coefficient matrix is now said to be an **upper triangular matrix**, since the only non-zero elements that it contains are on or above the **leading diagonal**, i.e. the diagonal from top left to bottom right, here containing the numbers 1, 10 and 2.

The leading diagonal is sometimes called the *main diagonal*.

Performing the matrix multiplication on the left-hand side gives the linear equations

$$\begin{aligned} x_1 - 4x_2 + 2x_3 &= -9, \\ 10x_2 - 3x_3 &= 34, \\ 2x_3 &= 4. \end{aligned}$$

Note that the upper triangular structure of the coefficient matrix means that each successive equation has one fewer unknown than its predecessor. Thanks to our successive elimination of unknown variables, we can see from the last equation that $x_3 = 2$, and we can substitute that into the equation immediately above it to find that $x_2 = 4$. This procedure is called **back substitution**. Having found the values of x_2 and x_3, we can perform another back substitution to find $x_1 = 3$.

So we have determined our solution $x_1 = 3$, $x_2 = 4$, $x_3 = 2$. However, it is always good practice to check it by writing down the solution in matrix form, and then using matrix multiplication to confirm that it solves the original matrix equation.

In this case $\mathbf{x} = [3 \quad 4 \quad 2]^T$, and we can check its correctness as follows:

$$\begin{bmatrix} 1 & -4 & 2 \\ 3 & -2 & 3 \\ 8 & -2 & 9 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 1\times 3 - 4\times 4 + 2\times 2 \\ 3\times 3 - 2\times 4 + 3\times 2 \\ 8\times 3 - 2\times 4 + 9\times 2 \end{bmatrix} = \begin{bmatrix} -9 \\ 7 \\ 34 \end{bmatrix}.$$

So our solution is correct. Even more pleasingly, it has been arrived at in a systematic way that can provide the basis of a general method for solving systems of simultaneous linear equations. This is the subject of the next subsection.

1.3 Gaussian elimination: non-singular cases

The technique that we have just been using is an example of *Gaussian elimination* – a general method for finding the unique solution of a system of linear equations, when such a solution exists. This subsection provides a more formal introduction to that method. As you will soon see, there are cases where no solution exists, other cases where an infinite number of solutions exist, and some cases where the solution is unique but the system of equations must be reformulated before Gaussian elimination can be used to find it. In this subsection and the one that follows, we will use examples and exercises to explore each of these exceptional cases, but first we set out a general procedure that is enough by itself to solve most cases of practical interest.

Remember that performing a row operation really involves three sub-steps:
- write down the plan
- implement the plan
- relabel the rows.

Procedure 1 Gaussian elimination

To solve a system of n linear equations in n unknowns, with coefficient matrix $\mathbf{A}$ and right-hand-side vector $\mathbf{b}$, it is often sufficient to carry out the following three steps.

1. **Formulation**: Write down the augmented matrix $\mathbf{A}|\mathbf{b}$ with rows $\mathbf{R}_1, \ldots, \mathbf{R}_n$.
2. **Elimination**: Adapt the following row operations as necessary.
   (a) Subtract a multiple of $\mathbf{R}_1$ from $\mathbf{R}_2$, to reduce to zero the first element in the first column below the leading diagonal.
   (b) Similarly, subtract a multiple of $\mathbf{R}_1$ from $\mathbf{R}_3, \ldots, \mathbf{R}_n$ to reduce to zero all the other elements in the first column below the leading diagonal.
   (c) In the new matrix obtained, subtract multiples of $\mathbf{R}_2$ from $\mathbf{R}_3, \ldots, \mathbf{R}_n$ to reduce to zero all the elements in the second column below the leading diagonal.
   (d) Continue this process until $\mathbf{A}|\mathbf{b}$ is reduced to $\mathbf{U}|\mathbf{c}$, where $\mathbf{U}$ is an upper triangular matrix.
3. **Solution**: Solve the system of equations with coefficient matrix $\mathbf{U}$ and right-hand-side vector $\mathbf{c}$ by back substitution.

Though not part of the procedure, it is generally good practice to write the solution as a column matrix, and then confirm by matrix multiplication that it is indeed a solution of the original system of equations.

Example 2

Solve the simultaneous equations

$$\begin{aligned} 3x_1 + x_2 - x_3 &= 1, \\ 5x_1 + x_2 + 2x_3 &= 6, \\ 4x_1 - 2x_2 - 3x_3 &= 3, \end{aligned}$$

and check the solution by matrix multiplication.

Solution

Following Procedure 1, starting with the formulation, the augmented matrix representing these equations is

$$\left[\begin{array}{ccc|c} 3 & 1 & -1 & 1 \\ 5 & 1 & 2 & 6 \\ 4 & -2 & -3 & 3 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array}.$$

Moving on to the elimination stage, to reduce the elements below the leading diagonal in column 1 to zero, replace $\mathbf{R}_2$ by $\mathbf{R}_2 - \frac{5}{3}\mathbf{R}_1$ and call the result $\mathbf{R}_{2a}$, then replace $\mathbf{R}_3$ by $\mathbf{R}_3 - \frac{4}{3}\mathbf{R}_1$ and call the result $\mathbf{R}_{3a}$:

It would be neater to replace $\mathbf{R}_2$ by $3\mathbf{R}_2 - 5\mathbf{R}_1$, and $\mathbf{R}_3$ by $3\mathbf{R}_3 - 4\mathbf{R}_1$, since this would avoid the use of fractions. We do not do this because it would involve combining elementary row operations where we prefer to use them individually at this stage. Nonetheless, such shortcuts are common, and we will use them later in the unit. You will probably use them yourself as your confidence grows.

$$\begin{array}{r} \\ \mathbf{R}_2 - \frac{5}{3}\mathbf{R}_1 \\ \mathbf{R}_3 - \frac{4}{3}\mathbf{R}_1 \end{array} \left[\begin{array}{ccc|c} 3 & 1 & -1 & 1 \\ 0 & -\frac{2}{3} & \frac{11}{3} & \frac{13}{3} \\ 0 & -\frac{10}{3} & -\frac{5}{3} & \frac{5}{3} \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array}.$$

To reduce the element below the leading diagonal in column 2 to zero, replace $\mathbf{R}_{3a}$ by $\mathbf{R}_{3a} - 5\mathbf{R}_{2a}$:

$$\begin{array}{r} \\ \\ \mathbf{R}_{3a} - 5\mathbf{R}_{2a} \end{array} \left[\begin{array}{ccc|c} 3 & 1 & -1 & 1 \\ 0 & -\frac{2}{3} & \frac{11}{3} & \frac{13}{3} \\ 0 & 0 & -20 & -20 \end{array}\right].$$

This completes the elimination stage. We now have to solve the equations represented by the new matrix, i.e.

$$\begin{aligned} 3x_1 + x_2 - x_3 &= 1, \\ -\tfrac{2}{3}x_2 + \tfrac{11}{3}x_3 &= \tfrac{13}{3}, \\ -20x_3 &= -20. \end{aligned}$$

It is clear from the last equation that $x_3 = 1$. So by back substitution, $x_2 = \frac{1}{2}(11x_3 - 13) = -1$ and $x_1 = \frac{1}{3}(1 - x_2 + x_3) = 1$.

We may verify that $\mathbf{x} = [1 \quad -1 \quad 1]^T$ is a solution of the matrix version of the equations as follows:

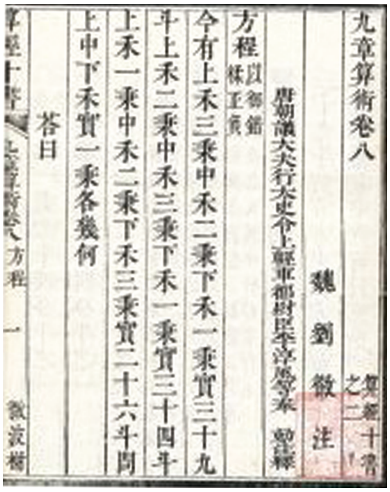
$$\begin{bmatrix} 3 & 1 & -1 \\ 5 & 1 & 2 \\ 4 & -2 & -3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3-1-1 \\ 5-1+2 \\ 4+2-3 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 3 \end{bmatrix}.$$

Exercise 3

Solve the simultaneous equations

$$
\begin{aligned}
x_1 + x_2 - x_3 &= 2,\\
5x_1 + 2x_2 + 2x_3 &= 20,\\
4x_1 - 2x_2 - 3x_3 &= 15,
\end{aligned}
$$

and check the solution by matrix multiplication.



九章算術卷八
魏 劉 徽 注
唐朝議大夫行太史令上輕車都尉臣李淳風等奉勅注釋
方程 以御錯糅正負
今有上禾三秉中禾二秉下禾一秉實三十九
斗上禾二秉中禾三秉下禾一秉實三十四斗
上禾一秉中禾二秉下禾三秉實二十六斗問
上中下禾實一秉各幾何
答曰

Figure 2 An ancient problem in linear algebra

All the examples of Gaussian elimination considered so far have started with a non-zero coefficient on the first variable in the first equation. One problem with Procedure 1 is that it will not work if that first entry in $\mathbf{R}_1$ is zero, since we would then have to divide by zero, which is not allowed. Fortunately, this is easy to overcome. All we have to do is reorder the equations in an appropriate way before formulating the augmented matrix. However, there is a deeper lesson to be learned from this. Reordering the equations is equivalent to interchanging complete rows in the augmented matrix, which is one of the allowed elementary row operations. Cases often arise in which it is advantageous to make such an interchange either before or even during the elimination stage. The following example illustrates this. (It is based on a problem set and solved in the Chinese text *Nine Chapters on the Mathematical Art* (Figure 2) written between 200 BC and 100 BC.)

Example 3

There are three types of corn, of which three bundles of the first, two of the second, and one of the third make 39 measures. Two of the first, three of the second and one of the third make 34 measures. And one of the first, two of the second and three of the third make 26 measures. How many measures of corn are contained in one bundle of each type?

Solution

Let x_i represent the number of measures in one bundle of type i, with $i = 1, 2, 3$. The problem is then specified by the simultaneous equations

$$
\begin{aligned}
3x_1 + 2x_2 + x_3 &= 39,\\
2x_1 + 3x_2 + x_3 &= 34,\\
x_1 + 2x_2 + 3x_3 &= 26.
\end{aligned}
$$

If we just press ahead and solve this problem as given, we will have to introduce fractions as soon as we try to eliminate x_1 from the second row. We can avoid the immediate use of fractions if we first interchange the first and third equations, so that the coefficient of x_1 in the first equation is 1. The augmented matrix of the reordered set of equations is then

$$
\left[\begin{array}{ccc|c}
1 & 2 & 3 & 26\\
2 & 3 & 1 & 34\\
3 & 2 & 1 & 39
\end{array}\right]
\begin{array}{l}
\mathbf{R}_1\\
\mathbf{R}_2\\
\mathbf{R}_3
\end{array}.
$$

Reducing the coefficients below the leading diagonal in the first column to zero gives

$$\begin{array}{r} \\ \mathbf{R}_2 - 2\mathbf{R}_1 \\ \mathbf{R}_3 - 3\mathbf{R}_1 \end{array} \left[\begin{array}{rrr|r} 1 & 2 & 3 & 26 \\ 0 & -1 & -5 & -18 \\ 0 & -4 & -8 & -39 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array}.$$

Reducing the coefficient below the leading diagonal in the second column to zero produces

$$\begin{array}{r} \\ \\ \mathbf{R}_{3a} - 4\mathbf{R}_{2a} \end{array} \left[\begin{array}{rrr|r} 1 & 2 & 3 & 26 \\ 0 & -1 & -5 & -18 \\ 0 & 0 & 12 & 33 \end{array}\right].$$

This completes the elimination. Back substitution then gives $x_3 = \frac{33}{12} = \frac{11}{4}$, $x_2 = 18 - 5x_3 = \frac{17}{4}$ and $x_1 = 26 - 2x_2 - 3x_3 = \frac{37}{4}$.

Now check that the solution $\mathbf{x} = \frac{1}{4}[37 \quad 17 \quad 11]^T$ satisfies the original matrix equation:

$$\frac{1}{4}\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 2 & 1 \end{bmatrix}\begin{bmatrix} 37 \\ 17 \\ 11 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 37 + 34 + 33 \\ 74 + 51 + 11 \\ 111 + 34 + 11 \end{bmatrix} = \begin{bmatrix} 26 \\ 34 \\ 39 \end{bmatrix}.$$

You may find it convenient to rearrange the data in the following exercise.

Exercise 4

The economy of Ruritania has collapsed, and each of its three banks has been nationalised. Under state ownership, all bank employees are divided into three grades: managers, software engineers and clerks, with members of a given grade paid identically, without bonuses. Bank A employs three managers, two software engineers and 24 clerks. Bank B employs one manager, one software engineer and 26 clerks. Bank C employs no managers, three software engineers and 25 clerks. Under this radically new dispensation, the monthly salary bill of each bank is 137 thousand rurs, where the rur is the name for the unit of the (recently devalued) currency. How many thousand rurs per month are received by a manager, by a software engineer and by a clerk?

In some cases it may be necessary to interchange a pair of rows later in the process, in order to achieve an upper triangular matrix at the end of the elimination stage. Here is an example where this happens.

Example 4

Solve the following system of equations:

$$\begin{aligned} x_1 + 10x_2 - 3x_3 &= 8, \\ x_1 + 10x_2 + 2x_3 &= 13, \\ x_1 + 4x_2 + 2x_3 &= 7. \end{aligned}$$

Solution

The corresponding augmented matrix is

$$\left[\begin{array}{ccc|c} 1 & 10 & -3 & 8 \\ 1 & 10 & 2 & 13 \\ 1 & 4 & 2 & 7 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array}.$$

To reduce the elements below the leading diagonal in column 1 to zero, replace $\mathbf{R}_2$ by $\mathbf{R}_2 - \mathbf{R}_1$ and call the result $\mathbf{R}_{2a}$, then replace $\mathbf{R}_3$ by $\mathbf{R}_3 - \mathbf{R}_1$ and call the result $\mathbf{R}_{3a}$:

$$\begin{array}{r} \\ \mathbf{R}_2 - \mathbf{R}_1 \\ \mathbf{R}_3 - \mathbf{R}_1 \end{array} \left[\begin{array}{ccc|c} 1 & 10 & -3 & 8 \\ 0 & 0 & 5 & 5 \\ 0 & -6 & 5 & -1 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array}.$$

In this case there is no need to perform another subtraction because a zero has fortuitously appeared in $\mathbf{R}_{2a}$. All that is needed is an interchange of the last two rows:

$$\begin{array}{r} \\ \mathbf{R}_{3a} \\ \mathbf{R}_{2a} \end{array} \left[\begin{array}{ccc|c} 1 & 10 & -3 & 8 \\ 0 & -6 & 5 & -1 \\ 0 & 0 & 5 & 5 \end{array}\right].$$

It follows that $x_3 = 1$, and by back substitution $x_2 = -\frac{1}{6}(-1 - 5x_3) = 1$ and $x_1 = 8 - 10x_2 + 3x_3 = 1$. (As usual, the matrix solution $\mathbf{x} = [1 \quad 1 \quad 1]^T$ should be confirmed as correct, as a check.)

In the previous example, the final exchange of rows was made to achieve an upper triangular matrix. However, it would have been possible to have solved the system of equations as soon as the unexpected zero appeared in $\mathbf{R}_{2a}$. In each of the following exercises, a solution becomes possible as soon as x_1 is eliminated from all but one of the equations, even though an upper triangular matrix is not apparent. It is left to you to decide whether to make use of this potential shortcut or to follow the systematic approach adopted earlier.

Exercise 5

Solve the system of equations whose augmented matrix is

$$\left[\begin{array}{ccc|c} 0 & 0 & 1 & 2 \\ 2 & 1 & 0 & 5 \\ 1 & 2 & 0 & 7 \end{array}\right].$$

Exercise 6

Solve the system of equations whose augmented matrix is

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 10 \\ 1 & 1 & 2 & 1 & 12 \\ 1 & 3 & 1 & 1 & 16 \\ 4 & 1 & 1 & 1 & 22 \end{array}\right].$$

A comment on the efficiency of Gaussian elimination

All of the examples and exercises in this subsection have involved a square coefficient matrix $\mathbf{A}$ with $\det \mathbf{A} \neq 0$. Such matrices are described as **non-singular** and, as explained in Unit 4, are invertible, meaning that it is possible to construct an inverse matrix, denoted $\mathbf{A}^{-1}$, such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. Such matrices may be contrasted with *singular* matrices that have $\det \mathbf{A} = 0$ and are non-invertible. The non-singular case is important because of the following general result.

The system of linear equations represented by the matrix equation $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is a square matrix, has a unique solution if and only if $\mathbf{A}$ is non-singular (i.e. $\det \mathbf{A} \neq 0$).

Such a system of equations is said to be a **non-singular system**.

In the non-singular case, the unique solution of the system of equations $\mathbf{Ax} = \mathbf{b}$ is given by $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, where $\mathbf{A}^{-1}$ is the *inverse* of $\mathbf{A}$, as defined in Unit 4.

It can be shown that the method of Gaussian elimination will always produce this solution, provided that we include any necessary interchanges of rows. Nonetheless, you may wonder why we have chosen to introduce Gaussian elimination when we could have used the methods of Unit 4 to construct the inverse of the coefficient matrix, and then used matrix multiplication to find the solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. The reason concerns the *efficiency* of the calculation, which becomes an increasingly important consideration as the number of equations grows and the order of the coefficient matrix increases.

The method for finding a matrix inverse that was described in the previous unit involved a lot of manipulation. Using this method to find the inverse of the $n \times n$ coefficient matrix $\mathbf{A}$ might be sensible if n is small, e.g. $n = 3$ or $n = 4$, but is a bad idea when n is large, for at least two reasons.

- The method given in Unit 4 takes of order $n!$ multiplications to compute $\mathbf{A}^{-1}$. Other methods, based on row operations, take roughly n^3 arithmetic operations. Since n^3 increases far more slowly than $n!$, as n gets large, the methods based on row operations are much more efficient.

$5! = 120$ and $6! = 720$, while $5^3 = 125$ and $6^3 = 216$.

- Even if we use a faster method to compute $\mathbf{A}^{-1}$, we may find that we require high-precision multiplication to achieve only modest accuracy in our result, since there may be large cancellations between the many terms summed over in our calculations.

For these (and other) reasons, the methods of Unit 4 are not those that are implemented by software packages that solve linear systems of equations. Gaussian elimination is widely regarded as a more efficient method for solving medium to large systems of linear equations – but there are exceptions to this, as the following box indicates.

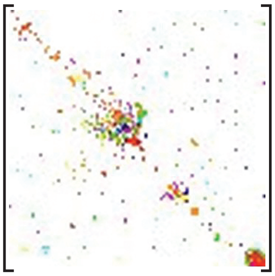


Figure 3 A sparse matrix of order 100×100. The value of each element is indicated by the colour of the corresponding pixel (i.e. picture element). By far the most common value is zero, which is indicated by white.

Gaussian elimination and sparse matrices

A large matrix $\mathbf{A}$ in which many of the elements are zero is said to be a *sparse matrix*. Such matrices arise in various applications, ranging from the modelling of engineering structures and electrical networks, to pure mathematical studies in the field known as number theory. An example of a sparse matrix is shown in Figure 3. In such cases, Gaussian elimination may be an inefficient method for solving $\mathbf{Ax} = \mathbf{b}$, since the development of an upper triangular matrix may well create many more non-zero elements than were initially present. Special methods have been devised for dealing with sparse matrices, which can handle millions of linear equations.

To end this discussion of Gaussian elimination in non-singular cases, we note that, despite its general efficiency, Gaussian elimination should not be thoughtlessly applied to every system of linear equations. In particular, you should not apply it to a system consisting of just two equations, since such a small system is more easily analysed without the formal apparatus of Gaussian elimination.

1.4 Gaussian elimination: singular cases

If $\det \mathbf{A} = 0$, then we say that $\mathbf{A}$ is *singular*, and we also use that term to describe the corresponding system of linear equations $\mathbf{Ax} = \mathbf{b}$. In some singular cases there may be no solution; in others there may be an infinity of solutions. Some 2×2 examples will help us to see why.

By this stage it is assumed that you can envisage the relevant 2×2 coefficient matrix without needing to write it down. If this is not the case, treat this example as a further exercise.

Example 5

For each of the following pairs of simultaneous equations, write down the determinant of the corresponding coefficient matrix and determine whether there is no solution, a unique solution, or an infinity of solutions.

(a) $x + y = 4,\quad 2x - 3y = 8.$ (b) $x + y = 4,\quad 2x + 2y = 5.$ (c) $x + y = 4,\quad 2x + 2y = 8.$

Solution

(a) In this case the determinant of the coefficient matrix is $-3 - 2 = -5$, so the system is non-singular and there is a unique solution.

(b) The determinant of the coefficient matrix is $2 - 2 = 0$, so the system is singular. Subtracting twice the first equation from the second, we obtain $0 = -3$, which is impossible. Hence there is no solution. (The equations are said to be **inconsistent**.)

(c) The determinant of the coefficient matrix is again $2 - 2 = 0$, so the system is singular. Subtracting twice the first equation from the second, we now obtain $0 = 0$, which is indeed true. The equations are consistent in this case. In fact, the second equation is simply the first equation multiplied throughout by 2. Consequently, any pair of values

that satisfies the first equation will also satisfy the second equation. There are infinitely many such pairs (choose any value for x and then let $y = 4 - x$). Hence there is an infinity of solutions.

Note that in the example above, the evaluation of the determinant was sufficient to tell us whether a unique solution existed. However, in the singular cases, where there was no unique solution, the value of the determinant did not tell us whether there was no solution or an infinity of solutions. Further investigation is always needed to resolve such cases. This is true irrespective of the order of $\mathbf{A}$. We can gain some geometrical insight into these different behaviours by recalling that a linear equation in two variables can be represented graphically by a straight line. We generally write the equation of such a line as $y = mx + c$, where m represents the *gradient* of the line, and c represents the *intersection* with the y-axis.

The lines corresponding to the equations of Example 5(a) are depicted in Figure 4(a), and are seen to intersect at a unique point. The values of x and y at the intersection represent the unique solution to the linear equations. The lines represented by the equations of Example 5(b) are parallel (see Figure 4(b)), so they never meet and there is no pair of values that simultaneously satisfies the two equations. Finally, in Figure 4(c), the two lines, though superficially different, are actually the same, so any of the infinity of points on one line will also be on the other. Hence there is an infinity of solutions.

A linear equation in three variables (e.g. $ax + by + cz = d$) represents a plane in a three-dimensional system of Cartesian coordinates. The constants (a, b, c and d) will determine the orientation of the plane and its perpendicular distance from the origin (the plane passes through the origin if $d = 0$). A system of three linear equations with a non-singular coefficient matrix of order 3×3 describes three planes that intersect at a unique point. The coordinates of this point represent the unique solution of the non-singular system of equations. This is illustrated in Figure 5(a).

Figures 5(b) and 5(c) show cases in which the planes meet in two or three parallel lines. These are some of the cases where the determinant of the coefficient matrix will be zero, and the singular system of equations will have no solution. Another singular case in which there is no solution is illustrated in Figure 5(d), where the three planes are parallel and do not intersect at all.

In Figure 5(e) we see a case where there are infinitely many solutions since all three planes share a common line. (The case of all three planes coinciding would also represent an infinity of solutions.) In these cases the coefficient matrix is again singular.

This geometrical interpretation can be extended to higher dimensions using the idea of a *hyperplane*, but visualisation is increasingly difficult so we will not pursue that approach here. In any case, the basic point has already been well established and can be summarised as follows.

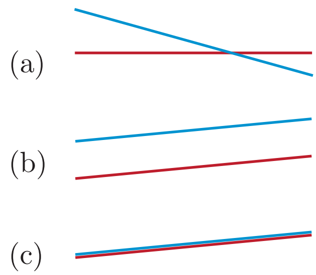


Figure 4 In two dimensions, a pair of simultaneous equations in two variables represents a pair of straight lines. Such lines may (a) intersect at a unique point, (b) be parallel and not intersect at all, or (c) be identical and intersect at an infinity of points.

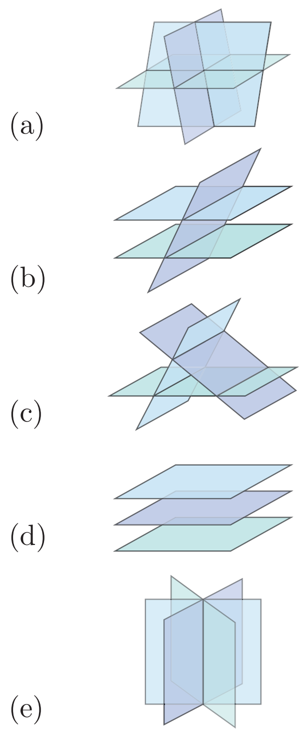


Figure 5 In three dimensions, the planes that represent three linear equations may (a) intersect at a unique point, (b) intersect in two parallel lines, (c) intersect in three parallel lines, (d) not intersect at all, or (e) intersect in the same line or plane

The system of linear equations represented by the matrix equation $\mathbf{Ax} = \mathbf{b}$ is said to be **singular** when the square matrix $\mathbf{A}$ is singular (i.e. when $\det \mathbf{A} = 0$). Such a system does not possess a unique solution; it may have no solution or an infinity of solutions.

The following exercise is important. It should help you to recognise the possible cases, particularly those in which there is an infinity of solutions.

In the case of singular systems, this exercise requires you to go no further than considering the *number* of solutions that a system will have. Examples and exercises that involve determining such solutions will be given in Section 3.

Exercise 7

For each of the following systems of simultaneous equations, determine whether there is no solution, a unique solution, or an infinity of solutions. (*Hint*: Examining determinants will not be sufficient; you are advised to consider the augmented matrices.)

(a)
$$\begin{aligned} x_1 - 2x_2 + 5x_3 &= 7, \\ x_1 + 3x_2 - 4x_3 &= 20, \\ x_1 + 18x_2 - 31x_3 &= 40. \end{aligned}$$

(b)
$$\begin{aligned} x_1 - 2x_2 + 5x_3 &= 6, \\ x_1 + 3x_2 - 4x_3 &= 7, \\ 2x_1 + 6x_2 - 12x_3 &= 12. \end{aligned}$$

(c)
$$\begin{aligned} x_1 - 4x_2 + x_3 &= 14, \\ 5x_1 - x_2 - x_3 &= 2, \\ 6x_1 + 14x_2 - 6x_3 &= -52. \end{aligned}$$

Figure 6 Carl Friedrich Gauss (1777–1855)

Gaussian elimination – a historical perspective

Carl Friedrich Gauss (Figure 6) is widely regarded as one of the greatest mathematicians of all time, comparable with Archimedes and Newton. He spent most of his career as Professor of Astronomy and Director of the Observatory at the University of Göttingen, in Lower Saxony, now part of Germany. One of his many celebrated achievements was the determination of the orbit of the asteroid Pallas from a small number of observations in 1802. In 1810 he published an important paper about orbit determinations in which he used repeated observations of Pallas to make the best possible determination of its orbit. During this work he took a systematic approach to finding the six unknowns in the system of equations that represented the observations. In doing so, he devised a notation that became widely adopted. It is from this that the method of Gaussian elimination gets its name.

Interestingly, the current – rather broad – use of the term 'Gaussian elimination' may not be entirely justified, even if we ignore the evidence that priority belongs to the Chinese. It appears that the name of Gauss became attached to the general method of systematic elimination only in the 1950s, mainly as a result of confusion among mathematicians who were unaware of the method's detailed history. The original inventor (at least as far as Europe is concerned) seems to have been Isaac Newton (Figure 7), who pointed out in 1670 that algebra textbooks lacked a procedure for solving systems of simultaneous linear equations, and proceeded to supply one.

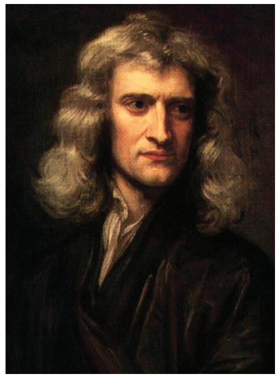


Figure 7 Isaac Newton (1642–1727)

2 Introducing eigenvalues and eigenvectors

As mentioned in the Introduction, eigenvalues and eigenvectors are scalars (often numbers) and vectors associated with a matrix (the German word 'eigen' may be translated as 'inherent'). They are widely used in the mathematical sciences, particularly physics. Rather than going directly to the definitions, we will first examine some examples of how they arise. We start with what is essentially an algebraic example, based on a mathematical model of population growth. This should help to establish a clear view of an eigenvector. Then, in a second example that is more geometrically based, we give particular emphasis to the idea of an eigenvalue.

'Eigen' is pronounced as 'eye-ghen'.

2.1 Eigenvectors in a model of population growth

Consider the (fictional) towns of Exton and Wyeville, which have a regular interchange of population: each year, one-tenth of Exton's population migrates to Wyeville, while one-fifth of Wyeville's population migrates to Exton (see Figure 8). Other changes in population, such as births, deaths and other migrations, cancel each other and so can be ignored. If x_n represents the population of Exton at the *beginning* of year n, and y_n is the corresponding population of Wyeville, then the populations of the two towns at the beginning of year $n+1$ are given by

$$x_{n+1} = 0.9x_n + 0.2y_n,$$
$$y_{n+1} = 0.1x_n + 0.8y_n,$$

which may be expressed in matrix form as

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix}.$$

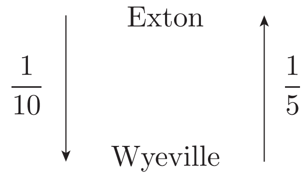


Figure 8 Exton and Wyeville annually exchange fixed fractions of their respective populations

This matrix equation can be represented symbolically by

$$\mathbf{x}_{n+1} = \mathbf{T}\mathbf{x}_n,$$

where the column vectors $\mathbf{x}_n$ and $\mathbf{x}_{n+1}$ represent the populations in Exton and Wyeville at the beginning of years n and $n+1$, respectively, and the square matrix $\mathbf{T}$ is known as the **transition matrix** for the problem. (The entries in such a matrix are all non-negative, and the entries in each column sum to 1.)

This annual exchange of populations is an example of an *iterative process*, in which the values associated with the $(n+1)$th iterate can be determined from the values associated with the nth iterate.

Suppose that initially the population of Exton is 10 000 and that of Wyeville is 8000, i.e. $\mathbf{x}_0 = [x_0 \quad y_0]^T = [10\,000 \quad 8000]^T$. Then after one year (i.e. at the beginning of year $n = 1$), the populations will be given by $\mathbf{x}_1 = \mathbf{T}\mathbf{x}_0$, so

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} 10\,000 \\ 8000 \end{bmatrix} = \begin{bmatrix} 10\,600 \\ 7400 \end{bmatrix},$$

and after two years they are given by $\mathbf{x}_2 = \mathbf{T}\mathbf{x}_1$, so

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} 10\,600 \\ 7400 \end{bmatrix} = \begin{bmatrix} 11\,020 \\ 6980 \end{bmatrix}.$$

Note that we might also have written this last result as

$$\mathbf{x}_2 = \mathbf{T}\mathbf{x}_1 = \mathbf{T}\mathbf{T}\mathbf{x}_0 = \mathbf{T}^2\mathbf{x}_0,$$

where we have introduced the power notation $\mathbf{T}^2$ to indicate repeated (matrix) multiplication, just as we use powers such as a^2, a^3 and a^4 to indicate the repeated multiplication of a scalar a. Using the idea of the **power of a matrix** to represent repeated matrix multiplication, we can describe the populations of Exton and Wyeville after n years as

$$\mathbf{x}_n = \mathbf{T}^n\mathbf{x}_0. \qquad (2)$$

Given this relation, how do the populations of Exton and Wyeville change over time? The answers can be worked out using equation (2), though the repeated matrix multiplications are tedious to do by hand. To save you the labour, the results are indicated graphically in Figure 9.

As you can see, the populations show a very interesting kind of behaviour. Initially they diverge, changing significantly from year to year, with Exton growing while Wyeville shrinks. However, those changes diminish with time, and after about 15 years the populations change very little, approaching ever more closely 12 000 in Exton and 6000 in Wyeville.

Explicit calculations (with results displayed to only the nearest integer) show that $\mathbf{x}_{29} = \mathbf{x}_{30} = \mathbf{x}_{31} = [12\,000 \quad 6000]^T$. So after about 30 years, the two populations are very stable, despite continuing annual migrations.

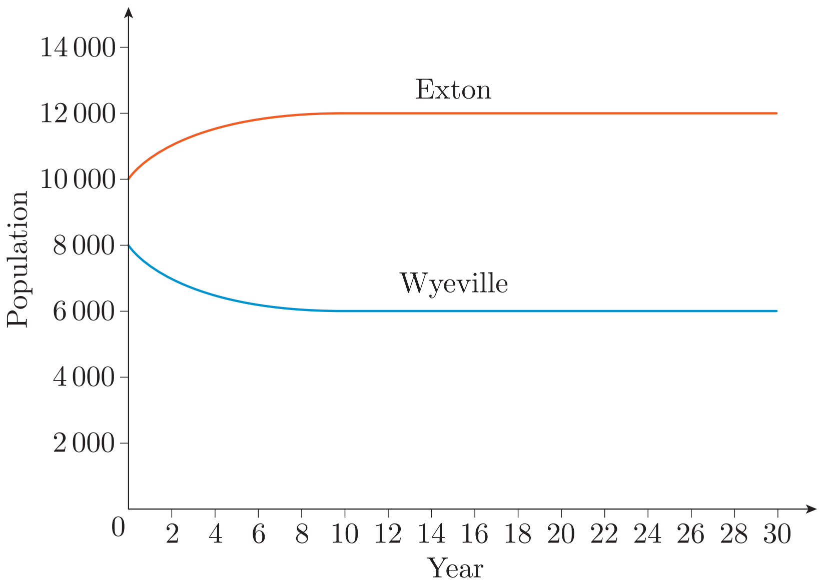


Figure 9 The evolving populations of Exton and Wyeville

Clearly there is something special about the relationship between the 'steady-state' population vector $\mathbf{x} = [x \quad y]^T = [12\,000 \quad 6000]^T$ and the transition matrix $\mathbf{T}$ for this particular problem. It is easy to see the nature of that special relationship by simply working out the matrix product of $\mathbf{T}$ and $\mathbf{x}$:

$$\begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} 12\,000 \\ 6000 \end{bmatrix} = \begin{bmatrix} 0.9 \times 12\,000 + 0.2 \times 6000 \\ 0.1 \times 12\,000 + 0.8 \times 6000 \end{bmatrix} = \begin{bmatrix} 12\,000 \\ 6000 \end{bmatrix}.$$

Expressed in symbolic terms,

$$\mathbf{Tx} = \mathbf{x},$$

showing that the action of $\mathbf{T}$ on $\mathbf{x}$ leaves $\mathbf{x}$ unchanged. This obviously explains the stability of the steady-state populations, since repeated applications of the transition matrix will continue to produce the same outcome. This special relationship between $\mathbf{x}$ and $\mathbf{T}$ is described by saying that $\mathbf{x}$ is an *eigenvector* of $\mathbf{T}$.

In fact, this particular example of an eigenvector is very special indeed. Rather more common are cases where $\mathbf{A}$ is a square matrix (not necessarily a transition matrix) and $\mathbf{y}$ is a column vector such that

$$\mathbf{Ay} = \lambda\mathbf{y},$$

where λ is a constant scalar. In these cases we would still describe $\mathbf{y}$ as an *eigenvector* of $\mathbf{A}$, but we would now say that λ is the corresponding *eigenvalue*. In the case of the population model considered above, the eigenvalue λ is 1, because $\mathbf{Tx} = 1\mathbf{x}$. In the next subsection we will consider another example of an eigenvector, but this time we will be specifically concerned with situations in which the corresponding eigenvalue is not 1.

Exercise 8

Use matrix multiplication to determine which, if any, of the following column vectors is an eigenvector of the transition matrix

$$\mathbf{T} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix}.$$

(a) $\begin{bmatrix} 1000 \\ 1000 \end{bmatrix}$ (b) $\begin{bmatrix} 120 \\ 60 \end{bmatrix}$ (c) $\begin{bmatrix} 500 \\ 300 \end{bmatrix}$ (d) $\begin{bmatrix} 20 \\ 10 \end{bmatrix}$

It is clear from this exercise that an eigenvector of a square matrix is not unique. In fact, given any eigenvector $\mathbf{x}$ of a square matrix, any scaled column vector $\mathbf{y} = k\mathbf{x}$, where k is a non-zero scalar constant, will also be an eigenvector of that square matrix. Moreover, $\mathbf{x}$ and $\mathbf{y}$ will correspond to the same eigenvalue λ. (It is important not to confuse the scalar multiple k discussed here with the eigenvalue λ that we will discuss in the next subsection.)

2.2 Eigenvalues in a transformation of the plane

In Unit 4 we saw that a 2×2 matrix $\mathbf{A} = [a_{ij}]$ can be interpreted geometrically as representing a linear transformation of the Cartesian plane. Such transformations map a point with the Cartesian coordinates (x_0, y_0) to a point with the Cartesian coordinates (x_1, y_1), where $x_1 = a_{11}x_0 + a_{12}y_0$ and $y_1 = a_{21}x_0 + a_{22}y_0$. In terms of matrices we can write this as

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}.$$

Geometrically, we may regard the coordinates that appear in the column vector on the right as the components of a position vector, $\mathbf{r}_0 = (x_0, y_0) = x_0\mathbf{i} + y_0\mathbf{j}$, where $\mathbf{i}$ and $\mathbf{j}$ are the unit vectors in the x- and y-directions, respectively. Thus, in geometric terms, the action of the matrix $\mathbf{A}$ is to transform or 'map' one vector $\mathbf{r}_0$ into another vector $\mathbf{r}_1$.

With this geometric view in mind, consider the linear transformation specified by the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}.$$

Using matrix multiplication, it is easy to see that this particular transformation will map the unit vector $\mathbf{i} = (1, 0)$ to the vector $(3, 1)$, and the unit vector $\mathbf{j} = (0, 1)$ to the vector $(2, 4)$. (If this is not clear, you should explicitly work out the matrix products $\mathbf{Ai}$ and $\mathbf{Aj}$, with $\mathbf{i}$ and $\mathbf{j}$ interpreted as column vectors.) These transformations are illustrated in Figure 10, along with the effect of the transformation on a general vector $\mathbf{r} = (x, y)$ that is mapped to the vector $(3x + 2y, x + 4y)$.

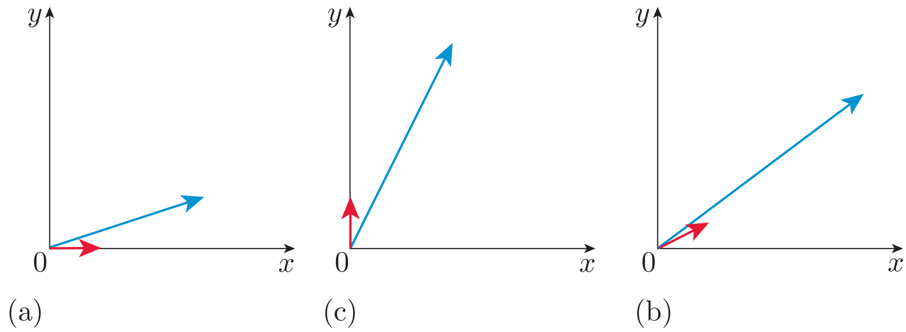


Figure 10 The action of the linear transformation represented by $\mathbf{A}$ on (a) the unit vector $\mathbf{i}$, (b) the unit vector $\mathbf{j}$, and (c) the general vector $\mathbf{r} = (x, y)$. In each case the initial vector is red and the transformed vector is blue.

Note that the action of the linear transformation on a general vector is to change its direction and its magnitude.

Now consider the action of $\mathbf{A}$ on the column vector $\mathbf{w} = [1 \quad 1]^T$:

$$\mathbf{A}\mathbf{w} = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}.$$

According to the rules of matrix algebra, we can write this last result as

$$\mathbf{A}\mathbf{w} = \begin{bmatrix} 5 \\ 5 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 5\mathbf{w}.$$

This shows that $\mathbf{w}$ is an *eigenvector* of $\mathbf{A}$, and that it corresponds to an *eigenvalue* of 5. The geometric interpretation of this result is shown in Figure 11; the transformation maps an eigenvector $\mathbf{w}$ into the scaled vector $5\mathbf{w}$, preserving its direction but altering its magnitude.

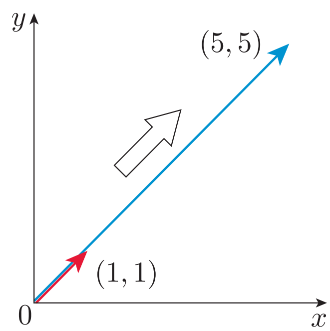


Figure 11 $\mathbf{A}$ maps its eigenvector $\mathbf{w}$ into the scaled vector $5\mathbf{w}$

Exercise 9

Use matrix multiplication to confirm that the linear transformation represented by $\mathbf{A}$ maps the scaled vector $k\mathbf{w}$, where k is any non-zero scalar, into the vector $5k\mathbf{w}$, and comment on the significance of this result.

The vector $\mathbf{w}$ and its scaled versions $k\mathbf{w}$ are not the only eigenvectors of $\mathbf{A}$, nor is 5 the only eigenvalue. Using matrix multiplication, it is also easy to show that $\mathbf{z} = [-2 \quad 1]^T$ is an eigenvector, and that in this case the corresponding eigenvalue is 2:

$$\mathbf{A}\mathbf{z} = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -4 \\ 2 \end{bmatrix} = 2\mathbf{z}.$$

This means, of course, that any non-zero scaled vector of the form $k\mathbf{z}$ will also be an eigenvector of $\mathbf{A}$, corresponding to the eigenvalue 2. This is why we describe $\mathbf{z} = [-2 \quad 1]^T$ as *an* eigenvector corresponding to eigenvalue 2, rather than *the* eigenvector corresponding to that eigenvalue. (You might like to confirm for yourself that $[4 \quad -2]^T$ and $[-1 \quad \frac{1}{2}]^T$ are also eigenvectors that correspond to the eigenvalue 2, since they are each of the form $k\mathbf{z}$, with $k = -1$ and $k = \frac{1}{4}$, respectively.)

The vectors $\mathbf{w}$ and $\mathbf{z}$ together with their scaled variants comprise all the possible eigenvectors of $\mathbf{A}$, and the eigenvalues 2 and 5 are the only eigenvalues of $\mathbf{A}$. So the particular transformation that we have been examining has two distinct eigenvalues, both of which happen to be positive. It should be noted, however, that this is not always the case. Eigenvalues may be positive or negative or zero, and real, imaginary or complex. As you will see later, a matrix may even have repeated eigenvalues, so while each eigenvector corresponds to a single eigenvalue, there is no guarantee that each eigenvector corresponds to a *different* eigenvalue.

Exercise 10

In each of the following cases, verify that $\mathbf{v}$ is an eigenvector of $\mathbf{A}$, and write down the corresponding eigenvalue.

(a) $\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$

(b) $\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$

(c) $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}.$

(d) $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$

Exercise 11

The term **real matrix** means that the matrix has real elements.

The real 2×2 matrix $\begin{bmatrix} 3 & -2 \\ 4 & -1 \end{bmatrix}$ has the complex eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1-i \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1+i \end{bmatrix}.$$

Show that the eigenvalues of the matrix are $1-2i$ and $1+2i$, and determine which eigenvalue corresponds to which eigenvector.

Sometimes it is possible to deduce information about the eigenvectors and eigenvalues of a given matrix from its geometric effects. This is so for each of the three cases considered in Figure 12. In each case, the geometric action of the matrix is illustrated by its effect on a unit square with vertices $(0,0)$, $(1,0)$, $(1,1)$ and $(0,1)$. Also shown is the effect of the transformation on the perpendicular unit vectors drawn along two adjoining sides of the unit square.

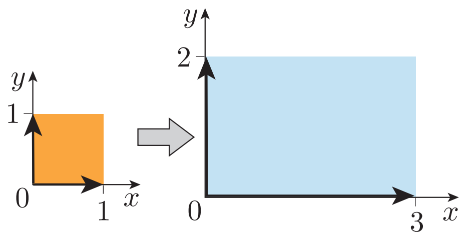
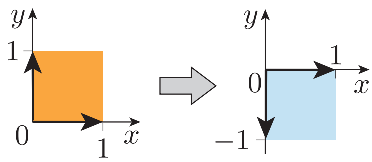
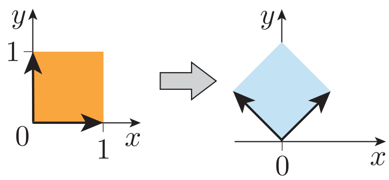
Matrix	Comment	Transformation of the unit square	Eigenvectors	Eigenvalues
$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$	A scaling by 3 in the x-direction and by 2 in the y-direction (i.e. a $(3,2)$ scaling)		$[1 \quad 0]^T$ $[0 \quad 1]^T$	3 2
$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	A reflection in the x-axis		$[1 \quad 0]^T$ $[0 \quad 1]^T$	1 -1
$\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$	A rotation through $\frac{\pi}{4}$ anticlockwise about the origin		No real eigenvectors	No real eigenvalues

Figure 12 Three matrices representing transformations of the plane, together with their real eigenvectors and corresponding eigenvalues

In the case of the matrix $\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$, it is clear from the geometric properties of the linear transformation that vectors along the coordinate axes $\mathbf{i} = [1 \quad 0]^T$ and $\mathbf{j} = [0 \quad 1]^T$ are eigenvectors, as these are transformed to vectors in the same directions. Vector $\mathbf{i}$ (and its scalar multiples) has eigenvalue 3, and vector $\mathbf{j}$ (and its scalar multiples) has eigenvalue 2.

In the case of the matrix $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, we see that reflection in the x-axis leaves the vector represented by $\mathbf{i} = [1 \quad 0]^T$ unchanged, and reverses the direction of $\mathbf{j} = [0 \quad 1]^T$, so these must be eigenvectors corresponding to the eigenvalues 1 and -1, respectively.

In the third case of Figure 12, the matrix describes rotation through $\pi/4$ anticlockwise about the origin. In this case we would not expect to find any real eigenvectors because the direction of every vector is changed by the linear transformation. However, even this matrix does have eigenvectors and eigenvalues; they simply happen to involve complex quantities. This puts them beyond the kind of simple geometric interpretation that we are using here, though they can be studied using the algebraic methods that will be introduced in Section 3.

Exercise 12

The matrix $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ corresponds to reflection in a line through the origin at an angle $\pi/4$ to the x-axis. What are the eigenvectors of $\mathbf{A}$ and their corresponding eigenvalues?

(*Hint*: Find two lines through the origin that are transformed to themselves by the reflection, then consider what happens to a point on each line.)

2.3 The eigenvalue equation

Having seen several examples, we are now in a good position to write down some formal definitions and list some of the properties of eigenvectors and eigenvalues. We begin with the simple but very important relationship often referred to as the *eigenvalue equation*.

The restriction to non-zero column vectors means that we never have to deal with an eigenvector of the form $\mathbf{v} = [0 \quad \cdots \quad 0]^T$.

Eigenvalue equation

Let $\mathbf{A}$ be any square matrix. A non-zero column vector $\mathbf{v}$ is an **eigenvector** of $\mathbf{A}$ if it satisfies the **eigenvalue equation** for $\mathbf{A}$:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \quad \text{for some scalar } \lambda. \tag{3}$$

The scalar λ is said to be the **eigenvalue** of $\mathbf{A}$ that corresponds to the eigenvector $\mathbf{v}$.

Exercise 13

Confirm that the transition matrix (considered in Subsection 2.1)

$$\mathbf{T} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix}$$

has eigenvectors $[2 \quad 1]^T$ and $[1 \quad -1]^T$, and find the corresponding eigenvalues.

If we count all the scaled versions of an eigenvector as equivalent, it can be shown that no $n \times n$ matrix may have more than n eigenvectors and n eigenvalues. Thus the two eigenvalues of $\mathbf{T}$ quoted in Exercise 13 are the *only* eigenvalues of $\mathbf{T}$, and the two given eigenvectors are its only eigenvectors (apart from their scalar multiples).

When dealing with a general $n \times n$ matrix, there are several complications that can arise when counting the number of eigenvalues and eigenvectors. We will say something about those complications in Section 3. For the present, however, we avoid them by considering only the simplest and most common case, in which an $n \times n$ matrix has n distinct eigenvalues and n distinct eigenvectors. This is what many would regard as the 'normal'

case, so it is natural to give it particular attention. Even so, we need to be clear about what we mean by 'distinct' eigenvalues and 'distinct' eigenvectors.

As far as eigenvalues are concerned, the notion of 'distinct' is simple: it just means different, i.e. unequal. The situation regarding eigenvectors is not so straightforward. We have already indicated that to be regarded as distinct, an eigenvector must not be a scalar multiple of any other eigenvector. However, with future developments in mind, we really need to go beyond this to define what we mean by distinct eigenvectors. In fact, we require a new concept called *linear independence*, which is the topic of the next subsection.

2.4 Linear independence

For the moment, let us forget about eigenvectors and just think about vectors in general. Linear independence is a property of sets of vectors. We begin with a simple geometric approach, first for two vectors, then for three vectors.

Two (non-zero) vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are said to be *linearly dependent* if they are collinear, i.e. one is parallel or antiparallel to the other. Conversely, $\mathbf{v}_1$ and $\mathbf{v}_2$ are said to be *linearly independent* if they are not collinear (see Figure 13).

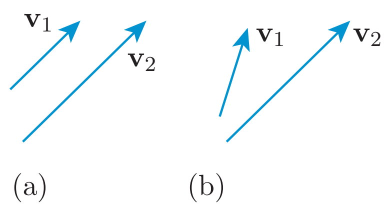


Figure 13 (a) Two linearly dependent vectors. (b) Two linearly independent vectors.

Now consider the case of three vectors. Three (non-zero) vectors $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ are said to be *linearly dependent* if they are coplanar, i.e. they all lie in the same plane. Conversely, $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ are said to be *linearly independent* if they are not coplanar (see Figure 14).

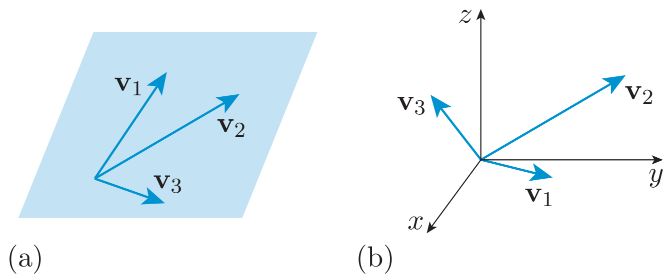


Figure 14 (a) A plane (in blue), containing three linearly dependent vectors. (b) Three linearly independent vectors in three-dimensional space.

Considering pairs and triples of vectors is as far as we can go using the geometric approach. So let us introduce an algebraic approach that will allow us to generalise to any number of vectors.

As stated, two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are *linearly dependent* if they are collinear, which is equivalent to $\mathbf{v}_1 = k\mathbf{v}_2$ for some number k. So they are *linearly independent* if $\mathbf{v}_1 \neq k\mathbf{v}_2$ for any k. Another way of saying this is: $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent if the only solution of the equation

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 = \mathbf{0} \quad \text{(where } \alpha_1 \text{ and } \alpha_2 \text{ are numbers)} \tag{4}$$

is $\alpha_1 = \alpha_2 = 0$.

This is equivalent to the previous statement because if we could find a solution of (4) with, say, $\alpha_1 \neq 0$, then $\mathbf{v}_1 = k\mathbf{v}_2$ with $k = -\alpha_2/\alpha_1$.

Now, three vectors $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ are *linearly dependent* if they are coplanar. This is equivalent to saying that one vector can be expressed as a linear combination of the other two, say $\mathbf{v}_1 = k_2\mathbf{v}_2 + k_3\mathbf{v}_3$ for some numbers k_2 and k_3 (for example, in Figure 14(a), the vectors satisfy $\mathbf{v}_2 = \mathbf{v}_1 + \mathbf{v}_3$). The vectors are *linearly independent* if this is *not* possible. Mathematically we state this as follows: $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ are linearly independent if the only solution of the equation

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \alpha_2\mathbf{v}_3 = \mathbf{0} \quad \text{(where the } \alpha_i \text{ are numbers)} \tag{5}$$

is $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

If we could find a solution of equation (5) with, say, $\alpha_1 \neq 0$, then we would have

$$\mathbf{v}_1 = -\frac{\alpha_2}{\alpha_1}\mathbf{v}_2 - \frac{\alpha_3}{\alpha_1}\mathbf{v}_3.$$

You should by now be noticing a pattern emerging, so we move to the general case. n vectors $\mathbf{v}_i$ $(i = 1, \ldots, n)$ are said to be linearly dependent if one of the vectors can be expressed as a linear combination of the others, e.g. $\mathbf{v}_1 = k_2\mathbf{v}_2 + k_3\mathbf{v}_3 + \cdots + k_n\mathbf{v}_n$ for some numbers k_i. If it is not possible to do this, then we say that the vectors are linearly independent. This is equivalent to the following definition.

Linear independence

n vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are **linearly independent** if the only solution of the equation

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0} \quad \text{(where the } \alpha_i \text{ are numbers)} \tag{6}$$

is $\alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_n = 0$.

If the vectors are not linearly independent, then they are said to be **linearly dependent**, and one of the vectors can be expressed as a linear combination of the others.

Example 6

In two dimensions, are the unit vectors $\mathbf{i} = [1 \quad 0]^T$ and $\mathbf{j} = [0 \quad 1]^T$ linearly dependent or linearly independent?

Solution

Clearly, $\alpha_1\mathbf{i} + \alpha_2\mathbf{j}$ never gives the zero vector unless $\alpha_1 = \alpha_2 = 0$. Hence the vectors are linearly independent.

This is also clear geometrically, since the vectors are neither parallel nor antiparallel.

Example 7

Show that the vectors $\mathbf{i} = [1 \quad 0 \quad 0]^T$, $\mathbf{j} = [0 \quad 1 \quad 0]^T$ and $\mathbf{v} = [3 \quad 2 \quad 0]^T$ are linearly dependent.

Solution

Since $\mathbf{v} = 3\mathbf{i} + 2\mathbf{j}$, we can write $\mathbf{v} - 3\mathbf{i} - 2\mathbf{j} = \mathbf{0}$. This is equivalent to equation (6), with $n = 3$, $\mathbf{v}_1 = \mathbf{v}$, $\mathbf{v}_2 = \mathbf{i}$, $\mathbf{v}_3 = \mathbf{j}$, $\alpha_1 = 1$, $\alpha_2 = -3$ and $\alpha_3 = -2$. So the vectors are linearly dependent.

This is also clear geometrically, since all three vectors lie in the xy-plane.

Exercise 14

Show that if there is a solution of equation (6) with $\alpha_1 \neq 0$, then one of the vectors can be written as a linear combination of the others.

Exercise 15

Are the following sets of three-dimensional column vectors linearly independent or linearly dependent? Justify your answers.

(a) $\mathbf{i} = [1 \quad 0 \quad 0]^T$, $\mathbf{j} = [0 \quad 1 \quad 0]^T$, $\mathbf{k} = [0 \quad 0 \quad 1]^T$.

(b) $\mathbf{v}_1 = [1 \quad -1 \quad 0]^T$, $\mathbf{v}_2 = [-1 \quad 1 \quad 0]^T$, $\mathbf{v}_3 = [0 \quad 0 \quad 1]^T$.

It is the concept of linear independence that really captures what we mean by ‘distinct’ eigenvectors: the eigenvectors of a matrix are said to be distinct if they are linearly independent. We return to this in the next subsection, but for now, let us continue our exploration of linear independence.

Linear independence and dimension

Linear independence of vectors is intimately connected with the dimension of the space in which they live. In fact, we often use the definition of linear independence to define what is meant by the *dimension* of a **vector space**. (The term *vector space* is often used to describe spaces of vectors in higher dimensions.) It should be fairly obvious that in two dimensions, we can have no more than two linearly independent vectors, because if we had three, they would all lie in the same plane and hence be linearly dependent. So in two dimensions, if we have two linearly independent vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, then we can express any other vector as a linear combination of them: $\mathbf{v} = \alpha\mathbf{v}_1 + \beta\mathbf{v}_2$ (because $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}$ are linearly dependent); see Figure 15.

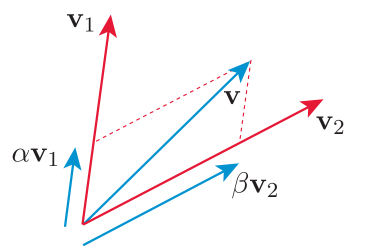


Figure 15 In two dimensions, any vector $\mathbf{v}$ may be written as a linear combination of any two linearly independent vectors $\mathbf{v}_1$ and $\mathbf{v}_2$

Likewise, in three dimensions we can have no more than three linearly independent vectors. And if we have three linearly independent vectors $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$, then we can express any other vector as a linear combination of them: $\mathbf{v} = \alpha\mathbf{v}_1 + \beta\mathbf{v}_2 + \gamma\mathbf{v}_3$ (because $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ and $\mathbf{v}$ are linearly dependent).

We use these observations to provide a definition of the dimension of a vector space.

This follows because in n dimensions, if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent and we add another vector $\mathbf{v}$, then $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly dependent, hence $\mathbf{v}$ can be expressed as a linear combination of the others.

Dimension and basis

The **dimension** of a vector space is equal to the maximum number of linearly independent vectors that it allows.

In an n-dimensional vector space, if we have n linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, then we can express any other vector as a linear combination of them:

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n. \tag{7}$$

The set of linearly independent vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is called a **basis** for the n-dimensional vector space.

Example 8

The vectors $\mathbf{v}_1 = [1 \quad 3]$ and $\mathbf{v}_2 = [2 \quad -1]$ are linearly independent. Express the vector $\mathbf{v} = [1 \quad 1]^T$ as a linear combination of them.

Solution

Setting $\mathbf{v} = \alpha\mathbf{v}_1 + \beta\mathbf{v}_2$, we have

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 3 \end{bmatrix} + \beta \begin{bmatrix} 2 \\ -1 \end{bmatrix},$$

from which we get the simultaneous linear equations

$$\begin{aligned} 1 &= \alpha + 2\beta, \\ 1 &= 3\alpha - \beta. \end{aligned}$$

Solving these, we obtain $\alpha = 3/7$ and $\beta = 2/7$.

Exercise 16

Are the following three-dimensional column vectors linearly independent or linearly dependent?

$$\mathbf{v}_1 = [1 \quad 0 \quad 2]^T, \quad \mathbf{v}_2 = [1 \quad -\tfrac{1}{2} \quad 0]^T,$$
$$\mathbf{v}_3 = [0 \quad \tfrac{1}{2} \quad -1]^T, \quad \mathbf{v}_4 = [0 \quad 0 \quad -2]^T.$$

Exercise 17

The vectors $\mathbf{v}_1 = [1 \quad 1]^T$ and $\mathbf{v}_2 = [-2 \quad 1]^T$ are linearly independent. Express the vector $\mathbf{v} = [1 \quad 3]^T$ as a linear combination of them.

Until now, in two dimensions we have taken the (linearly independent) unit Cartesian vectors $\mathbf{i}$ and $\mathbf{j}$ as the basis, and expressed every other vector as a linear combination of them: $\mathbf{v} = v_x\mathbf{i} + v_y\mathbf{j}$. Similarly, in three dimensions we have taken $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ as the basis, and expressed every other vector as a linear combination of them: $\mathbf{v} = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k}$.

However, what we now know is that it is not necessary to use orthogonal unit vectors as a basis. In n dimensions, *any* set of n vectors will suffice as a basis, provided that they are all linearly independent.

Having explored the notion of linear independence, we now return to our main topic and apply these ideas (which apply to vectors in general) to the eigenvectors of a matrix.

2.5 Application of linear independence to eigenvector expansions

Recall that given an $n \times n$ matrix $\mathbf{A}$, there are a set of numbers λ (the eigenvalues) and a set of vectors $\mathbf{v}$ (the eigenvectors) that satisfy the eigenvalue equation

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}.$$

Furthermore, if $\mathbf{v}$ is an eigenvector (with eigenvalue λ), then so is the scalar multiple $k\mathbf{v}$ (for any number k), so $\mathbf{v}$ and $k\mathbf{v}$ are not regarded as distinct eigenvectors.

In the 'usual' case there are n distinct eigenvalues with n distinct eigenvectors. 'Distinct eigenvalues' means that they are not equal. We are now in a position to say what we mean by the term 'distinct eigenvectors'.

A matrix has n distinct eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ if they are linearly independent.

Let us check that the 2×2 matrix discussed in Subsection 2.2 does indeed have only two linearly independent eigenvectors.

Example 9

In Subsection 2.2 we showed that the matrix $\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$ has eigenvectors of the form $[k \quad k]^T$, corresponding to the eigenvalue 5, and eigenvectors of the form $[-2k \quad k]^T$, corresponding to the eigenvalue 2. Answer the following questions regarding these eigenvectors, justifying each of your answers.

(a) Are the eigenvectors $[1 \quad 1]^T$ and $[2 \quad 2]^T$ linearly dependent?

(These correspond to the first eigenvector, with $k = 1$ and $k = 2$, respectively.)

(b) Are the eigenvectors $[-2 \quad 1]^T$ and $[2 \quad -1]^T$ linearly dependent?

(These correspond to the second eigenvector, with $k = 1$ and $k = -1$, respectively.)

(c) Are the eigenvectors $[1 \quad 1]^T$ and $[-2 \quad 1]^T$ linearly dependent?

(These correspond to the first and second eigenvectors, both with $k = 1$.)

Solution

(a) Yes, since

$$2[1 \quad 1]^T - [2 \quad 2]^T = [0 \quad 0]^T,$$

so the eigenvectors fail the test for linear independence, and are therefore linearly dependent. This is also clear geometrically, since the vectors are collinear.

(b) Yes, since

$$[-2 \quad 1]^T - [2 \quad -1]^T = [0 \quad 0]^T,$$

so the eigenvectors fail the test for linear independence, and are therefore linearly dependent. This is also clear geometrically, since the vectors are collinear.

(c) No. For the vectors to be linearly dependent, we would need to be able to find non-zero scalars α_1 and α_2 such that

$$\alpha_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

For this to be possible, we require

$$\begin{aligned} \alpha_1 - 2\alpha_2 &= 0, \\ \alpha_1 + \alpha_2 &= 0. \end{aligned}$$

However, the only solution of this pair of equations is $\alpha_1 = 0$ and $\alpha_2 = 0$, so the two eigenvectors corresponding to different eigenvalues are linearly independent (and therefore distinct).

This example demonstrates that although $\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$ has an infinity of eigenvectors of the forms $[k \quad k]^T$ and $[-2k \quad k]^T$, there are only two linearly independent eigenvectors. We are free to choose any value of k. Choosing $k = 1$, we say that $\mathbf{A}$ has two linearly independent eigenvectors $\mathbf{v}_1 = [1 \quad 1]^T$ and $\mathbf{v}_2 = [-2 \quad 1]^T$. In fact, since one only ever talks about *linearly independent* eigenvectors, a statement such as this is often abbreviated to '$\mathbf{A}$ has two eigenvectors, $\mathbf{v}_1 = [1 \quad 1]^T$ and $\mathbf{v}_2 = [-2 \quad 1]^T$'.

The obvious question is: when does an $n \times n$ matrix have n linearly independent eigenvectors? It turns out that *if the eigenvalues are distinct, then so are their corresponding eigenvectors.*

This is a powerful statement, but its proof, although not difficult, is beyond the scope of this module.

If an $n \times n$ matrix has n distinct eigenvalues, then it has n linearly independent eigenvectors.

In fact, in many (but not all) cases, even if the eigenvalues are not all distinct, we can still have n linearly independent eigenvectors.

Exercise 18

The matrix

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & 4 \\ -2 & 7 & -10 \\ -1 & 4 & -6 \end{bmatrix}$$

has distinct eigenvalues $\lambda_1 = -1$, $\lambda_2 = 1$, $\lambda_3 = 2$, and corresponding eigenvectors $\mathbf{v}_1 = [-1 \quad 1 \quad 1]^T$, $\mathbf{v}_2 = [1 \quad 2 \quad 1]^T$ and $\mathbf{v}_1 = [0 \quad 2 \quad 1]^T$ (up to a multiplicative constant). Show that the eigenvectors are linearly independent.

Eigenvector expansions

We have already seen that in an n-dimensional vector space, any set of n linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ can be used as a basis, i.e. any vector can be expressed as a linear combination of the $\mathbf{v}_i$ as in equation (7).

This means that in the 'normal' case that we are considering, when an $n \times n$ matrix $\mathbf{A}$ has n linearly independent eigenvectors, we can use these as the basis vectors. This is called an *eigenvector expansion*.

Eigenvector expansion

If $\mathbf{A}$ is an $n \times n$ matrix with n linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, and $\mathbf{v}$ is any n-dimensional vector, then there exist scalars $c_1, c_2, \dots, c_n$ such that

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n. \tag{8}$$

This is called the **eigenvector expansion** of $\mathbf{v}$.

Example 10

We saw in Exercise 13 that in the population model of Subsection 2.1, the transition matrix

$$\mathbf{T} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix}$$

has eigenvectors $[2 \quad 1]^T$ and $[1 \quad -1]^T$ that correspond to the distinct eigenvalues 1 and 0.7. Use this information to determine an eigenvector expansion of the two-dimensional column vector $\mathbf{x}_0 = [10\,000 \quad 8000]^T$ that represents the initial populations of Exton and Wyeville.

Solution

The eigenvectors correspond to different eigenvalues, so are linearly independent. This is also obvious because they are not collinear.

They therefore form a basis for two-dimensional vectors, so we can write

$$\mathbf{x}_0 = \begin{bmatrix} 10\,000 \\ 8000 \end{bmatrix} = c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2c_1 + c_2 \\ c_1 - c_2 \end{bmatrix}.$$

Equating corresponding elements of the column vectors on the right and the left shows that

$$2c_1 + c_2 = 10\,000 \quad \text{and} \quad c_1 - c_2 = 8000.$$

Solving this simple system of linear equations, we see that $c_1 = 6000$ and $c_2 = -2000$. Thus

$$\mathbf{x}_0 = 6000 \begin{bmatrix} 2 \\ 1 \end{bmatrix} - 2000 \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Why is an eigenvector expansion of this kind particularly useful? The answer lies in the very simple effect that a matrix has on its own eigenvectors. To take a two-dimensional example, let λ_1 and λ_2 be the eigenvalues corresponding to the linearly independent eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$, so $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ for $i = 1, 2$. It then follows from the eigenvector expansion of any vector $\mathbf{v}$ that

$$\mathbf{A}\mathbf{v} = \mathbf{A}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2.$$

The value of this kind of simplification will be made clear in the next subsection. However, before that you can apply it for yourself in the following exercise.

Exercise 19

We saw in Subsection 2.2 that the transformation of the plane represented by the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$$

has eigenvectors $[1 \quad 1]^T$ and $[-2 \quad 1]^T$ that correspond to the distinct eigenvalues 5 and 2.

(a) Use this information to determine an eigenvector expansion of the position column vector $\mathbf{r}_0 = [-2 \quad 4]^T$.

(b) Use the eigenvector expansion to calculate $\mathbf{A}\mathbf{r}_0$ without using matrix multiplication.

2.6 Convergence towards an eigenvector

We can use what we have just learned about eigenvector expansions to provide insight into the behaviour of the population model of Subsection 2.1.

You will recall that the annual changes in population of Exton and Wyeville were represented by the action of a transition matrix $\mathbf{T}$ that transformed the populations at the beginning of year n into those at the beginning of year $n+1$. We already know, from Example 10, that the initial populations of Exton and Wyeville can be written as

$$\mathbf{x}_0 = \begin{bmatrix} 10\,000 \\ 8000 \end{bmatrix} = 6000\mathbf{v}_1 - 2000\mathbf{v}_2,$$

where $\mathbf{v}_1 = [2 \quad 1]^T$ and $\mathbf{v}_2 = [1 \quad -1]^T$ are the eigenvectors corresponding to the eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 0.7$ of the matrix $\mathbf{T}$.

Now consider the effect of repeatedly applying the transition matrix $\mathbf{T}$ to the initial population vector $\mathbf{x}_0 = 6000\mathbf{v}_1 - 2000\mathbf{v}_2$. Remembering that $\mathbf{T}\mathbf{v}_1 = \lambda_1\mathbf{v}_1$ and $\mathbf{T}\mathbf{v}_2 = \lambda_2\mathbf{v}_2$, a single application of $\mathbf{T}$ to $\mathbf{x}_0$ gives

$$\begin{aligned} \mathbf{x}_1 = \mathbf{T}\mathbf{x}_0 &= \mathbf{T}(6000\mathbf{v}_1 - 2000\mathbf{v}_2) \\ &= 6000\mathbf{T}\mathbf{v}_1 - 2000\mathbf{T}\mathbf{v}_2 \\ &= 6000\lambda_1\mathbf{v}_1 - 2000\lambda_2\mathbf{v}_2. \end{aligned}$$

A second application of $\mathbf{T}$ gives

$$\mathbf{x}_2 = \mathbf{T}\mathbf{x}_1 = \mathbf{T}(6000\lambda_1\mathbf{v}_1 - 2000\lambda_2\mathbf{v}_2) = 6000\lambda_1^2\mathbf{v}_1 - 2000\lambda_2^2\mathbf{v}_2.$$

A third application of $\mathbf{T}$ gives

$$\mathbf{x}_3 = \mathbf{T}\mathbf{x}_2 = \mathbf{T}(6000\lambda_1^2\mathbf{v}_1 - 2000\lambda_2^2\mathbf{v}_2) = 6000\lambda_1^3\mathbf{v}_1 - 2000\lambda_2^3\mathbf{v}_2,$$

and after k applications,

$$\mathbf{x}_k = \mathbf{T}\mathbf{x}_{k-1} = 6000\lambda_1^k\mathbf{v}_1 - 2000\lambda_2^k\mathbf{v}_2.$$

Now, $\lambda_1 = 1$, so λ_1^k is also equal to 1. However, $\lambda_2 = 0.7$, so $\lambda_2^2 = 0.49$, $\lambda_2^3 = 0.343$ and $\lambda_2^{30} = 0.000\,022\,5$ (to three significant figures). As we repeatedly apply $\mathbf{T}$ to $\mathbf{x}_0$, we will find that the contribution of $\mathbf{v}_2$ will become smaller and smaller as k increases, so we have

$$\mathbf{x}_k \simeq 6000\mathbf{v}_1 = \begin{bmatrix} 12\,000 \\ 6000 \end{bmatrix} \quad \text{for large } k.$$

This is exactly what happened in the population model, where we found that the populations approached 12 000 in Exton and 6000 in Wyeville.

Suppose that we start with some other initial population $\mathbf{x}_0$. Because we can always write $\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$, repeated application of $\mathbf{T}$ will give

$$\begin{aligned} \mathbf{x}_k &= \mathbf{T}^k(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) \\ &= c_1\mathbf{T}^k\mathbf{v}_1 + c_2\mathbf{T}^k\mathbf{v}_2 \\ &= c_1\lambda_1^k\mathbf{v}_1 + c_2\lambda_2^k\mathbf{v}_2 \\ &\simeq c_1\mathbf{v}_1 \quad \text{for large } k. \end{aligned}$$

More generally, suppose that we have an arbitrary $n \times n$ matrix $\mathbf{A}$ and a vector $\mathbf{x}$. What will be the result of repeated application of $\mathbf{A}$ to $\mathbf{x}$, i.e. what is $\mathbf{A}^k\mathbf{x}$? If $\mathbf{A}$ has n linearly independent eigenvectors $\mathbf{v}_i$, then we can always write $\mathbf{x}$ as an eigenvector expansion

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

Let us assume that we have chosen our initial vector $\mathbf{x}$ so that no c_i is zero. Then since $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$, where λ_i is the eigenvalue corresponding to eigenvector $\mathbf{v}_i$, we have

$$\mathbf{A}^k \mathbf{v}_i = \lambda_i^k \mathbf{v}_i.$$

Hence

$$\begin{aligned}\mathbf{A}^k \mathbf{x} &= \mathbf{A}^k (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n)\\ &= c_1\mathbf{A}^k\mathbf{v}_1 + c_2\mathbf{A}^k\mathbf{v}_2 + \cdots + c_n\mathbf{A}^k\mathbf{v}_n\\ &= c_1\lambda_1^k\mathbf{v}_1 + c_2\lambda_2^k\mathbf{v}_2 + \cdots + c_n\lambda_n^k\mathbf{v}_n.\end{aligned}$$

As k gets larger and larger, the eigenvalue with largest modulus will dominate. So if λ_p is the eigenvalue with largest modulus, we have

$$\mathbf{A}^k \mathbf{x} \simeq c_p \lambda_p^k \mathbf{v}_p \quad \text{for large } k.$$

So we see that for (almost) any initial vector $\mathbf{x}$, for large k, $\mathbf{A}^k\mathbf{x}$ is proportional to the eigenvector with largest modulus eigenvalue. And since we ignore any scaling of eigenvectors, we can say that $\mathbf{A}^k\mathbf{x}$ is equal to the eigenvector with largest modulus eigenvalue. This observation is the basis for many numerical algorithms for finding eigenvectors of very large matrices.

Exercise 20

Use $\mathbf{A}$, $\mathbf{r}_0$ and the eigenvector expansion of Exercise 19 to determine (to two significant figures) $\mathbf{r}_8 = \mathbf{A}^8\mathbf{r}_0$, $\mathbf{r}_9 = \mathbf{A}^9\mathbf{r}_0$ and $\mathbf{r}_{10} = \mathbf{A}^{10}\mathbf{r}_0$. Comment on your results.

Google's PageRank algorithm: the world's largest eigenvector problem

Before Google, web search engines were a hit and miss affair, often returning masses of irrelevant links. The advent of Google, in the late 1990s, changed all that. Apart from its phenomenal speed, it seemed to deliver 'the best' pages available to a given search. The principal reason for this is its PageRank algorithm (see Figure 16), named after its co-inventor Larry Page (Figure 17), who also co-founded Google.

Figure 16 An illustration of how the PageRank algorithm works: the size of each face is proportional to the total size of the other faces that are pointing to it

Figure 17 Larry Page, co-inventor of the PageRank algorithm

The PageRank algorithm quantitatively assesses the 'importance' of each page on the web by the number of other 'important' web pages that link to it. One way to get a feel for how it works is to imagine playing a (rather dull) game. Start on any web page and click on any link at random. From the new web page, again click on any link at random. Keep doing this, very many times, randomly going from web page to web page via the links. The PageRank of a particular web page is the proportion of time spent visiting that web page.

Mathematically, this is modelled as follows. First label every page on the web by an integer i (the order is not important). Then construct a huge matrix $\mathbf{A}$, where every row and column represents a page on the web. ($\mathbf{A}$ is currently approximately of order $10^{10} \times 10^{10}$, corresponding to the $\simeq 10^{10}$ current web pages.) Roughly speaking, each element A_{ij} is set equal to the probability of going from web page i to web page j by randomly clicking on a link; e.g. $A_{5,23}$ is the probability of visiting web page 23 by randomly clicking on a link on web page 5. Now consider a vector, each element of which represents the probability of being on a web page. Starting the game on web page i is represented by the vector $\mathbf{x}_0$, all of whose elements are zero except for the ith element, which is unity. The vector $\mathbf{x}_1 = \mathbf{A}\mathbf{x}_0$ gives the probability of being on any web page after one click. The vector $\mathbf{x}_n = \mathbf{A}^n\mathbf{x}_0$ gives the probability of being on any web page after n clicks. The situation is very similar to the Exton–Wyeville population model, and the final PageRank is given by the eigenvector of $\mathbf{A}$ with largest eigenvalue, i.e. $\mathbf{x}_n$ for very large n. The web page with largest PageRank corresponds to the largest element of $\mathbf{x}_n$.

3 Finding eigenvalues and eigenvectors

You have now seen several examples of eigenvectors and eigenvalues, and some of their uses. You have also been told a little about finding eigenvectors and eigenvalues through iteration. The time has now come for a more detailed investigation of the determination of eigenvalues and eigenvectors. In the next subsection we introduce the important notion of the *characteristic equation* of a matrix. This is a polynomial equation, the roots of which are the eigenvalues of the matrix. Once the eigenvalues have been found, methods based on row operations can be used to construct the related eigenvectors. We will examine some examples based on 2×2 and 3×3 matrices, but you should be aware that many of the ideas are very general, and that modern applied mathematics makes extensive use of computer packages to find the eigenvectors and eigenvalues of matrices, often based on the row operations and iterative methods that have already been mentioned.

3.1 The characteristic equation

An eigenvector $\mathbf{v}$ of an $n \times n$ matrix $\mathbf{A}$ satisfies $\mathbf{Av} = \lambda\mathbf{v}$. By introducing an $n \times n$ identity matrix $\mathbf{I}$, this can be written as $\mathbf{Av} = \lambda\mathbf{Iv}$, i.e.

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}. \tag{9}$$

Clearly $\mathbf{v} = \mathbf{0}$ is a solution of this equation – called the **trivial solution**. We are not interested in this because, by definition, $\mathbf{v} = \mathbf{0}$ is not an eigenvector. For there to be non-trivial solutions $\mathbf{v}$, the $n \times n$ square matrix on the left, $(\mathbf{A} - \lambda\mathbf{I})$, known as the **characteristic matrix** of $\mathbf{A}$, must be non-invertible. (If it were invertible, the unique solution would be $\mathbf{v} = (\mathbf{A} - \lambda\mathbf{I})^{-1}\mathbf{0} = \mathbf{0}$, which is trivial.) For the characteristic matrix to be non-invertible, it must be singular, i.e. its determinant must be zero: $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. Expanding the determinant gives a polynomial equation of degree n satisfied by every eigenvalue of $\mathbf{A}$.

For example, in the case of a 2×2 matrix $\mathbf{A} = [a_{ij}]$ and an eigenvector $\mathbf{v} = [v_1 \quad v_2]^T$, equation (9) is

$$\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which can also be written as

$$\begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

So for a non-trivial solution, we require

$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = 0.$$

The determinant here gives a quadratic equation for λ, whose two solutions are the eigenvalues of $\mathbf{A}$.

Characteristic equation

Let $\mathbf{A}$ be any square matrix. The equation

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \tag{10}$$

is called the **characteristic equation** of $\mathbf{A}$. Its roots, i.e. the values of λ that satisfy the characteristic equation, are the eigenvalues of $\mathbf{A}$.

Example 11

Write out the characteristic equation of the transition matrix introduced in the Exton–Wyeville population model,

$$\mathbf{T} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix},$$

and use it to determine the two eigenvalues that arise in that model.

Solution

In this case, the characteristic equation is given by

$$\begin{vmatrix} 0.9-\lambda & 0.2 \\ 0.1 & 0.8-\lambda \end{vmatrix} = 0.$$

Expanding the determinant gives

$$(0.9-\lambda)(0.8-\lambda) - 0.1 \times 0.2 = 0,$$

which can be rewritten as

$$\lambda^2 - 1.7\lambda + 0.7 = 0.$$

This is a quadratic equation in λ that may be either solved by the standard formula or factorised to give

$$(\lambda - 1.0)(\lambda - 0.7) = 0.$$

By either approach, it is clear that there are only two roots, $\lambda_1 = 1.0$ and $\lambda_2 = 0.7$. You learned earlier that these are the two eigenvalues of $\mathbf{T}$.

Exercise 21

Write out the characteristic equation of the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix},$$

and hence find the eigenvalues of $\mathbf{A}$.

3.2 Eigenvalues of 2×2 matrices

The following procedure, based on the characteristic equation, can be used to find the two eigenvalues of any 2×2 matrix.

Procedure 2 Finding eigenvalues of a 2×2 matrix

Let $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. To find the eigenvalues of $\mathbf{A}$, do the following.

1. Write down the characteristic equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.
2. Expand this as
$$\begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = \lambda^2 - (a+d)\lambda + (ad-bc) = 0. \tag{11}$$
3. Solve this quadratic equation to find the two values of λ, which are the required eigenvalues.

Exercise 22

Calculate the eigenvalues of the following matrices.

(a) $\mathbf{G} = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix}$ (b) $\mathbf{H} = \begin{bmatrix} 2 & 2 \\ -1 & 2 \end{bmatrix}$

Exercise 23

Calculate the eigenvalues of the following two-dimensional rotation and scaling matrices.

(a) $\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ (b) $\mathbf{M} = \begin{bmatrix} l & 0 \\ 0 & k \end{bmatrix}$

For a 2×2 matrix we can always solve the quadratic equation that results from the characteristic equation and hence determine the two eigenvalues. Nonetheless, it is worth investigating the 2×2 case a little further because of the light that it can shed on a number of problems.

As a first step we introduce the quantity known as the *trace* of a matrix.

Trace of a matrix

Given any $n \times n$ matrix $\mathbf{A} = [a_{ij}]$, the **trace** of that matrix is denoted $\operatorname{tr}\mathbf{A}$ and is the sum of the elements on its leading diagonal:

$$\operatorname{tr}\mathbf{A} = a_{11} + a_{22} + \cdots + a_{nn}. \qquad (12)$$

It follows from this definition that in the case of a general 2×2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, its trace is given by $\operatorname{tr}\mathbf{A} = a + d$. However, for such a matrix the determinant is $\det\mathbf{A} = ad - bc$. It therefore follows from equation (11) that the expanded form of the characteristic equation of a 2×2 matrix can be written as follows.

Note that this formula is valid *only* for 2×2 matrices.

Characteristic equation of a 2×2 matrix

$$\lambda^2 - \operatorname{tr}\mathbf{A}\,\lambda + \det\mathbf{A} = 0. \qquad (13)$$

Applying the generic formula for finding the roots of a quadratic equation, we see that the two eigenvalues are given by

$$\lambda = \tfrac{1}{2}\left(\operatorname{tr}\mathbf{A} \pm \sqrt{D}\right), \quad \text{where } D = (\operatorname{tr}\mathbf{A})^2 - 4\det\mathbf{A}. \qquad (14)$$

The quantity D is known as the **discriminant**. For a 2×2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the discriminant is

$$D = (a+d)^2 - 4(ad - bc) = (a-d)^2 + 4bc.$$

Hence the two eigenvalues can be expressed as follows.

Eigenvalues of a 2×2 matrix

$$\lambda = \tfrac{1}{2}\left(a + d \pm \sqrt{(a-d)^2 + 4bc}\right). \tag{15}$$

For a real 2×2 matrix we thus have three cases:

- a pair of distinct real eigenvalues if $D > 0$
- repeated real eigenvalues if $D = 0$
- a pair of complex eigenvalues if $D < 0$.

The first and third of these cases correspond to the normal situation with two distinct eigenvalues. Note, however, that when $D < 0$, we have $\lambda = \frac{1}{2}\left(\operatorname{tr}\mathbf{A} \pm \sqrt{-|D|}\right) = \frac{1}{2}\left(\operatorname{tr}\mathbf{A} \pm i\sqrt{|D|}\right)$, so the eigenvalues are complex, with one eigenvalue being the complex conjugate of the other. (You saw instances of this in Exercises 22 and 23.) The second case, when $D = 0$, means that the characteristic equation has a repeated root.

Exercise 24

Show that the characteristic equation of the matrix $\mathbf{S} = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$, where s is a number, has a repeated root, and determine what it is.

There are two useful results that follow directly from equation (14). First, note that adding the two eigenvalues gives

$$\lambda_1 + \lambda_2 = \operatorname{tr}\mathbf{A}.$$

Second, note that multiplying the eigenvalues gives

$$\begin{aligned}\lambda_1\lambda_2 &= \tfrac{1}{2}\left(\operatorname{tr}\mathbf{A} + \sqrt{D}\right)\tfrac{1}{2}\left(\operatorname{tr}\mathbf{A} - \sqrt{D}\right)\\ &= \tfrac{1}{4}\left((\operatorname{tr}\mathbf{A})^2 - D\right)\\ &= \det\mathbf{A}.\end{aligned}$$

Although we have derived these results for a 2×2 matrix, they both turn out to be generally true for square matrices of any order. The two general results are as follows.

Trace and determinant rules for eigenvalues

For any $n \times n$ matrix:

- the trace is equal to the sum of all its eigenvalues
- the determinant is equal to the product of all its eigenvalues.

These results are frequently used to assist and check matrix calculations.

Exercise 25

Given that $\lambda_1 = 26.115$ and $\lambda_2 = -0.115$ are the eigenvalues of the matrix

$$\mathbf{A} = \begin{bmatrix} 11 & 12 \\ 14 & 15 \end{bmatrix},$$

confirm that they pass the trace and determinant checks (working to two decimal places).

3.3 Eigenvalues of larger matrices

Formulating and solving the characteristic equation of some 3×3 matrices is not too daunting, but beyond that the task becomes manually challenging. With the wide availability of computer packages, there is a tendency to turn rapidly to a machine that will either perform the necessary algebra or provide numerical estimates of the eigenvalues.

Example 12

Write down the characteristic equation of the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & 1 \\ -1 & 1 & 5 \end{bmatrix},$$

and given that one of its roots is 6, find the other two roots. Confirm that the sum of the roots gives the trace of $\mathbf{A}$, and the product of the roots is the determinant of $\mathbf{A}$.

Solution

In this case the characteristic equation of $\mathbf{A}$ (equation (10)) is given by

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} 3-\lambda & -1 & -1 \\ -1 & 3-\lambda & 1 \\ -1 & 1 & 5-\lambda \end{vmatrix} = 0.$$

Using Laplace's rule (see Unit 4) to expand the determinant in terms of the elements of the top row gives

$$(3-\lambda)[(3-\lambda)(5-\lambda)-1] - (-1)[-(5-\lambda)+1] - 1[-1+(3-\lambda)] = 0,$$

which may be rewritten as

$$\lambda^3 - 11\lambda^2 + 36\lambda - 36 = 0. \qquad (16)$$

Knowing that one of the roots of this equation is $\lambda = 6$, we can extract a factor $(\lambda - 6)$ to obtain

$$(\lambda - 6)(\lambda^2 - 5\lambda + 6) = 0.$$

(Factorising an equation like this is best done in bits. First, knowing that $\lambda = 6$ is a root, we write $(\lambda - 6)(a\lambda^2 + b\lambda + c)$ for some constants a, b, c. Then comparing with equation (16) immediately gives $a = 1$ and $c = 6$. Finally, expanding $(\lambda - 6)(\lambda^2 + b\lambda + 6)$ and comparing with equation (16) gives $b = -5$.)

Then, either by using the usual formula to factorise the quadratic function, or factorising by inspection, we can write

$$(\lambda - 2)(\lambda - 3)(\lambda - 6) = 0.$$

This shows that the three eigenvalues are 2, 3 and 6.

Adding the three eigenvalues gives $2 + 3 + 6 = 11$, which is also the sum of the leading diagonal elements, $\operatorname{tr} \mathbf{A} = 3 + 3 + 5$. This confirms that the trace of $\mathbf{A}$ is equal to the sum of its eigenvalues.

Similarly, the product of the eigenvalues is $2 \times 3 \times 6 = 36$, while using Laplace's rule to expand $\det \mathbf{A}$ gives

$$\det \mathbf{A} = 3(15 - 1) - (-1)(-5 + 1) - 1(-1 + 3) = 42 - 4 - 2 = 36.$$

This confirms that the determinant of $\mathbf{A}$ is equal to the product of its eigenvalues.

Exercise 26

Find the eigenvalues of the following matrices, given that in each case one of the eigenvalues is 2.

(a) $\mathbf{A} = \begin{bmatrix} 1 & -2 & 4 \\ -2 & 7 & -10 \\ -1 & 4 & -6 \end{bmatrix}$ (b) $\mathbf{A} = \begin{bmatrix} 4 & 7 & 6 \\ 6 & 5 & 6 \\ -8 & -10 & -10 \end{bmatrix}$

Exercise 27

Check that your answers to Exercise 26 satisfy the trace and determinant rules for eigenvalues.

3.4 General results on eigenvalues

In this subsection we list some general results regarding the eigenvalues of certain types of $n \times n$ matrices. We mainly prove the results for 2×2 matrices, but you should note that they are valid in general.

We already know the trace and determinant rules, which are valid for any square matrix. Now let us consider some rules that apply to special types of matrices.

Real matrices

Recall that a **real matrix** is one whose elements are all real. In Subsection 3.2, we proved that for a real 2×2 matrix, if the eigenvalues are complex, then they occur in complex conjugate pairs; i.e. for a real matrix, if λ is an eigenvalue, then so is $\overline{\lambda}$.

It is trivial to prove this for any real $n \times n$ matrix. If $\mathbf{A}$ is a real matrix and λ is a complex eigenvalue with corresponding eigenvector $\mathbf{v}$, then $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. So taking the complex conjugate of both sides, we get $\overline{\mathbf{A}\mathbf{v}} = \overline{\lambda\mathbf{v}}$,

which is equivalent to

$$\overline{\mathbf{A}}\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}.$$

But $\mathbf{A}$ is real, so $\overline{\mathbf{A}} = \mathbf{A}$, and we have

$$\mathbf{A}\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}.$$

Hence $\overline{\lambda}$ is also an eigenvalue of $\mathbf{A}$, with corresponding eigenvector $\overline{\mathbf{v}}$.

Exercise 28

The matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

has one eigenvalue $\lambda_1 = 1 + i$. Determine the other two eigenvalues without solving the characteristic equation.

Triangular matrices

A matrix is **triangular** if all the entries above (or below) the leading diagonal are 0, e.g.

A **diagonal** matrix is a special type of triangular matrix.

$$\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}, \begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}, \quad \begin{bmatrix} a & 0 \\ c & d \end{bmatrix}, \begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix}, \quad \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}.$$

(upper triangular) (lower triangular) (diagonal)

For a 2×2 triangular matrix $\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$ or $\begin{bmatrix} a & 0 \\ c & d \end{bmatrix}$, the characteristic equation is

$$(a - \lambda)(d - \lambda) = 0,$$

hence the eigenvalues are $\lambda = a$ and $\lambda = d$. Thus the eigenvalues of a triangular matrix are the diagonal entries. This is true for any $n \times n$ triangular matrix, and the proof is very similar to that given above.

Exercise 29

What are the eigenvalues of the matrix $\begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$?

Real symmetric matrices

A matrix is **symmetric** if it is equal to its transpose, i.e. $\mathbf{A} = \mathbf{A}^T$ – that is, the entries are symmetric about the leading diagonal, e.g.

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}, \quad \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix}.$$

For a *real* 2×2 symmetric matrix $\mathbf{A} = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$, the characteristic equation is (see equation (11))

$$\lambda^2 - (a+d)\lambda + (ad - b^2) = 0,$$

hence the eigenvalues are

$$\lambda = \tfrac{1}{2}\left(a + d \pm \sqrt{(a+d)^2 - 4(ad - b^2)}\right).$$

The term under the square root (i.e. the discriminant) is

$$(a+d)^2 - 4(ad - b^2) = (a^2 + 2ad + d^2) - 4ad + 4b^2 = (a-d)^2 + 4b^2,$$

which is the sum of two squares and therefore cannot be negative. It follows that the eigenvalues of a real symmetric matrix are real.

Note that if a matrix has real eigenvalues, it is not necessarily true that it is a real symmetric matrix.

There is a well-known proof of this result for $n \times n$ real symmetric matrices, which we include in Subsection 4.2, but you are not required to know it.

Exercise 30

Under what circumstances can a symmetric matrix $\begin{bmatrix} a & b \\ b & d \end{bmatrix}$ have a repeated eigenvalue?

Non-invertible matrices

A *non-invertible* matrix has determinant equal to 0 (see Unit 4). However, from the determinant rule, if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of an $n \times n$ *non-invertible* matrix $\mathbf{A}$, then

$$\lambda_1 \lambda_2 \ldots \lambda_n = \det \mathbf{A} = 0.$$

It follows that a matrix is non-invertible if and only if at least one of its eigenvalues is 0. Also, a matrix is invertible if and only if all its eigenvalues are non-zero.

We summarise what we have learned.

Properties of eigenvalues – summary

- The product of the eigenvalues of $\mathbf{A}$ is $\det \mathbf{A}$.
- The sum of the eigenvalues of $\mathbf{A}$ is $\operatorname{tr} \mathbf{A}$.
- The complex eigenvalues and corresponding eigenvectors of a real matrix occur in complex conjugate pairs.
- The eigenvalues of a triangular matrix are the diagonal entries.
- The eigenvalues of a real symmetric matrix are real.
- A matrix is non-invertible if and only if at least one of its eigenvalues is 0.

Exercise 31

Without solving the characteristic equation, what can you say about the eigenvalues of each of the following matrices?

(a) $\mathbf{A} = \begin{bmatrix} 67 & 72 \\ 72 & -17 \end{bmatrix}$ (b) $\mathbf{A} = \begin{bmatrix} 67 & 72 \\ 0 & -17 \end{bmatrix}$ (c) $\mathbf{A} = \begin{bmatrix} 288 & 72 \\ 72 & 18 \end{bmatrix}$

We now turn to eigenvectors.

3.5 The eigenvector equation

An eigenvector $\mathbf{v}$ of a general $n \times n$ real matrix $\mathbf{A}$ satisfies $\mathbf{Av} = \lambda\mathbf{v}$. As you saw in Subsection 3.1 (equation (9)), by introducing an $n \times n$ identity matrix $\mathbf{I}$, this condition can be written as the following **eigenvector equation**.

Eigenvector equation

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}. \qquad (17)$$

Earlier we were interested in the determinant of the left-hand side that led to the characteristic equation and the eigenvalues. Now we need to find the non-zero solutions of the linear equations themselves. This is most easily described in the context of 2×2 matrices, as in the next subsection.

3.6 Eigenvectors of 2×2 matrices

Consider a 2×2 matrix

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Suppose that we have solved the characteristic equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ to find the eigenvalues of this matrix. Now let us try to find the eigenvectors. Suppose that the unknown eigenvector $\mathbf{v} = [x \quad y]^T$ has a known eigenvalue λ. Then the eigenvector equation

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$$

gives

$$\begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which is equivalent to the equations

$$(a-\lambda)x + by = 0, \qquad (18)$$
$$cx + (d-\lambda)y = 0. \qquad (19)$$

Since we know a, b, c, d and λ, this is a pair of simultaneous linear equations that can be solved for x and y. We did this in Section 1. However, there is a subtlety: because $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, the system of equations is singular. Such systems of linear equations were discussed in

Subsection 1.4, where we discovered that they can have either no solution or an infinity of solutions. In fact, it turns out that the two equations (18) and (19) are really the same equation written in two different-looking ways. So we need to solve only one of them, as the other gives no information, thus we obtain an infinity of solutions (see Example 5(c) for a similar case).

To prove that the two equations are the same, use the fact that $\det(\mathbf{A} - \lambda\mathbf{I}) = (a-\lambda)(d-\lambda) - bc = 0$.

We should not be surprised by this, because the eigenvector equation determines the eigenvectors only up to an arbitrary scalar multiple. So when we solve equation (18) or (19) for x and y, we obtain a solution of the form $\mathbf{v} = k[x \quad y]^T$, for an arbitrary (non-zero) scalar k.

If we want, we can assign an arbitrary value to k, such as $k = 1$. It should then be understood that any (non-zero) scalar multiple is also an eigenvector. However we choose to handle it, having found the eigenvector corresponding to λ, our next step should be to use the other eigenvalue of $\mathbf{A}$ to find a second eigenvector.

Here is a numerical example of what we have just been describing.

Example 13

Find the eigenvalues and corresponding eigenvectors of

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 1 & -2 \end{bmatrix}.$$

Solution

The characteristic equation is

$$\begin{vmatrix} 1-\lambda & 4 \\ 1 & -2-\lambda \end{vmatrix} = 0.$$

Expanding gives $(1-\lambda)(-2-\lambda) - 4 = 0$, which simplifies to $\lambda^2 + \lambda - 6 = 0$. So the eigenvalues are $\lambda = 2$ and $\lambda = -3$.

Let $\mathbf{v} = [x \quad y]^T$ be an eigenvector. Then the eigenvector equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ is

$$\begin{bmatrix} 1-\lambda & 4 \\ 1 & -2-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which is equivalent to the system

$$\begin{aligned} (1-\lambda)x + \qquad\quad 4y &= 0, \\ x + (-2-\lambda)y &= 0. \end{aligned}$$

Now consider these for each of the eigenvalues in turn.

- For $\lambda = 2$, the eigenvector equations become

  $$-x + 4y = 0 \quad \text{and} \quad x - 4y = 0.$$

  Clearly these equations are the same, so we have the single equation $4y = x$. This single equation shows that if $x = k$, then $y = k/4$. So the eigenvector corresponding to $\lambda = 2$ has the general form $[k \quad k/4]^T$. Choosing $k = 4$ for convenience gives $[4 \quad 1]^T$ as an eigenvector corresponding to eigenvalue $\lambda = 2$. As usual, any non-zero scalar multiple is also an eigenvector.

Note that even though the two equations provide the same information, both are examined since this is a useful check.

- For $\lambda = -3$, the eigenvector equations become

  $$4x + 4y = 0 \quad \text{and} \quad x + y = 0,$$

  which are equivalent to the single equation $y = -x$. This single equation tells us that if $x = k$, then $y = -k$. So the eigenvector corresponding to $\lambda = -3$ has the general form $[k \quad -k]^T$; choosing $k = 1$ gives $[1 \quad -1]^T$.

Of course, it is always good practice to check that we have correctly determined our eigenvalues and eigenvectors by checking that they satisfy $\mathbf{Av} = \lambda\mathbf{v}$ explicitly.

Here, as a summary, is the general procedure.

Procedure 3 Finding the eigenvectors of a 2×2 matrix

Let $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. To find an eigenvector corresponding to the eigenvalue λ, do the following.

1. Write down the eigenvector equations

$$(a - \lambda)x + by = 0, \tag{20}$$
$$cx + (d - \lambda)y = 0. \tag{21}$$

2. These equations reduce to a single equation that is readily solved for x and y. The eigenvector is given by $\mathbf{v} = [x \quad y]^T$, with x and y replaced by their solved values. Any non-zero scalar multiple is also an eigenvector.
3. Repeat for the second eigenvalue to find the second eigenvector.

Exercise 32

Find the eigenvalues and eigenvectors of the following matrices.

(a) $\mathbf{A} = \begin{bmatrix} 8 & -5 \\ 10 & -7 \end{bmatrix}$ (b) $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

Exercise 33

Find the eigenvectors of

$$\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

when θ is not an integer multiple of π.

(*Hint*: The eigenvalues $\cos\theta \pm i\sin\theta = e^{\pm i\theta}$ were found in Exercise 23.)

3.7 Eigenvectors of larger matrices

In the general case, if $\mathbf{A}$ is an $n \times n$ matrix, there will be n eigenvalues. For each eigenvalue λ, the eigenvector equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ will provide n simultaneous equations, and we must solve these to determine the n components of the corresponding eigenvector. As before, because $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, the system of n equations is singular with an infinity of solutions, corresponding to the fact that any (non-zero) scalar multiple of an eigenvector is also an eigenvector. That means that in at least one case (and perhaps more), we will not be able to determine a unique value for a component of $\mathbf{v}$, and we will have to represent that component by an arbitrary non-zero scalar k. The other components can then be expressed in terms of that value k.

Often, the best way of solving the system of eigenvector equations is to use Gaussian elimination, as described in Section 1. The procedure is to express the eigenvector equations in terms of an augmented matrix, then use row operations to reduce the coefficient matrix to upper triangular form. The fact that the system is singular will result in at least one row of zeros in the augmented matrix, indicating that the corresponding unknown can be assigned the arbitrary value k. This row can be made the bottom row of the augmented matrix. All the other components can then be expressed in terms of k, using back substitution.

Here is a numerical example.

Example 14

Find the three eigenvectors of

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & 4 \\ -2 & 7 & -10 \\ -1 & 4 & -6 \end{bmatrix},$$

given that the eigenvalues are -1, 1 and 2.

Solution

The eigenvector equation for the given matrix, for $\mathbf{v} = [x_1 \quad x_2 \quad x_3]^T$, may be written

$$\begin{bmatrix} 1-\lambda & -2 & 4 \\ -2 & 7-\lambda & -10 \\ -1 & 4 & -6-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

This is equivalent to the system

$$\begin{aligned} (1-\lambda)x_1 - \quad 2x_2 + \quad 4x_3 &= 0, \\ -2x_1 + (7-\lambda)x_2 - \quad 10x_3 &= 0, \\ -x_1 + \quad 4x_2 + (-6-\lambda)x_3 &= 0. \end{aligned}$$

- For $\lambda = -1$, the augmented matrix is

$$\begin{array}{c} \left[\begin{array}{ccc|c} 2 & -2 & 4 & 0 \\ -2 & 8 & -10 & 0 \\ -1 & 4 & -5 & 0 \end{array}\right] \end{array} \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array}.$$

Reducing the elements below the leading diagonal in column 1 to zero:

$$\begin{array}{r} \\ \mathbf{R}_2 + \mathbf{R}_1 \\ \mathbf{R}_3 + \frac{1}{2}\mathbf{R}_1 \end{array} \left[\begin{array}{ccc|c} 2 & -2 & 4 & 0 \\ 0 & 6 & -6 & 0 \\ 0 & 3 & -3 & 0 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array}.$$

Reducing the element below the leading diagonal in column 2 to zero:

$$\begin{array}{r} \\ \\ \mathbf{R}_{3a} - \frac{1}{2}\mathbf{R}_{2a} \end{array} \left[\begin{array}{ccc|c} 2 & -2 & 4 & 0 \\ 0 & 6 & -6 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The final row of zeros (a result of the expected singular system) allows us to assign x_3 the value k, which is arbitrary apart from the requirement that the resulting eigenvector should be non-zero. Back substitution then tells us that $6x_2 = 6k$, so $x_2 = k$, and back substituting again gives $2x_1 - 2k + 4k = 0$, so $x_1 = -k$. Thus the general form of the eigenvector is $\mathbf{v} = k[-1 \quad 1 \quad 1]^T$, where k is an arbitrary non-zero value.

- For $\lambda = 1$, the augmented matrix is

$$\left[\begin{array}{ccc|c} 0 & -2 & 4 & 0 \\ -2 & 6 & -10 & 0 \\ -1 & 4 & -7 & 0 \end{array}\right].$$

In this case it will be helpful to interchange rows before doing anything else, so the starting arrangement will be

$$\left[\begin{array}{ccc|c} -1 & 4 & -7 & 0 \\ 0 & -2 & 4 & 0 \\ -2 & 6 & -10 & 0 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array}.$$

Completing the reduction of the elements below the leading diagonal in column 1 to zero:

$$\begin{array}{r} \\ \\ \mathbf{R}_3 - 2\mathbf{R}_1 \end{array} \left[\begin{array}{ccc|c} -1 & 4 & -7 & 0 \\ 0 & -2 & 4 & 0 \\ 0 & -2 & 4 & 0 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_{3a} \end{array}.$$

Reducing the element below the leading diagonal in column 2 to zero:

$$\begin{array}{r} \\ \\ \mathbf{R}_{3a} - \mathbf{R}_2 \end{array} \left[\begin{array}{ccc|c} -1 & 4 & -7 & 0 \\ 0 & -2 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The final row of zeros allows us to assign x_3 the arbitrary non-zero value k. Back substitution then tells us that $x_2 = 2k$, and back substituting again gives $x_1 = k$. Thus the general form of the eigenvector is $\mathbf{v} = k[1 \quad 2 \quad 1]^T$, where k is an arbitrary non-zero value.

- For $\lambda = 2$, the augmented matrix is

$$\left[\begin{array}{rrr|r} -1 & -2 & 4 & 0 \\ -2 & 5 & -10 & 0 \\ -1 & 4 & -8 & 0 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array}.$$

Reducing the elements below the leading diagonal in column 1 to zero:

$$\begin{array}{l} \\ \mathbf{R}_2 - 2\mathbf{R}_1 \\ \mathbf{R}_3 - \mathbf{R}_1 \end{array} \left[\begin{array}{rrr|r} -1 & -2 & 4 & 0 \\ 0 & 9 & -18 & 0 \\ 0 & 6 & -12 & 0 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array}.$$

Reducing the element below the leading diagonal in column 2 to zero:

$$\begin{array}{l} \\ \\ \mathbf{R}_{3a} - \frac{2}{3}\mathbf{R}_{2a} \end{array} \left[\begin{array}{rrr|r} -1 & -2 & 4 & 0 \\ 0 & 9 & -18 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The final row of zeros allows us to assign x_3 the arbitrary non-zero value k. Back substitution then tells us that $x_2 = 2k$, and back substituting again gives $x_1 = 0$. Thus the general form of the eigenvector is $\mathbf{v} = k[0 \quad 2 \quad 1]^T$, where k is an arbitrary non-zero value.

Here, as a summary, is the general procedure for finding eigenvectors.

Procedure 4 Finding eigenvectors in general

Let $\mathbf{A}$ be an $n \times n$ matrix. To find the eigenvectors of $\mathbf{A}$, first solve its characteristic equation, to find the eigenvalues. Then for each eigenvalue λ, solve the eigenvector equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ to find $\mathbf{v}$.

If λ is not a repeated eigenvalue, this will determine an eigenvector up to an arbitrary scalar multiple.

When eigenvalues are repeated, there can be a number of subtleties, which we do not cover in this module.

Exercise 34

Find the eigenvalues and corresponding eigenvectors of the following matrices.

(a) $\begin{bmatrix} 2 & 1 & -1 \\ 0 & -3 & 2 \\ 0 & 0 & 4 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 2 & 0 \\ -2 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Exercise 35

Find the three eigenvectors of

$$\mathbf{A} = \begin{bmatrix} 4 & 7 & 6 \\ 6 & 5 & 6 \\ -8 & -10 & -10 \end{bmatrix},$$

given that the eigenvalues are -2, -1 and 2.

3.8 Eigenvectors of real symmetric matrices

In Subsection 3.4 we defined a symmetric matrix as one that is equal to its own transpose. There we showed that *the eigenvalues of a real symmetric matrix are real.* Real symmetric matrices have several practical applications. For example, the matrix that describes the system of pipes and taps given in the Introduction is a real symmetric matrix. In this subsection we investigate the eigenvectors of real symmetric matrices.

Given that the eigenvalues of a real symmetric matrix are necessarily real, it follows that the eigenvectors of such a matrix can always be chosen to be real. This is because the eigenvector equation that can be used to determine the eigenvectors contains only real elements, and the process of solving that equation will not introduce any imaginary quantities. However, reality is not the only interesting issue concerning the eigenvectors of real symmetric matrices.

We say that the eigenvectors can be *chosen* to be real because having found an eigenvector, we can always choose to multiply it by an imaginary or complex number if we wish, and it will still be an eigenvector.

In Exercise 32(b) we showed that the real symmetric matrix $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ has the (real) eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 3$, and corresponding eigenvectors $\mathbf{v}_1 = [1 \quad -1]^T$ and $\mathbf{v}_2 = [1 \quad 1]^T$.

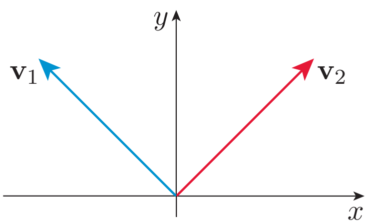


Figure 18 Eigenvectors of a real symmetric matrix corresponding to distinct (real) eigenvalues

These eigenvectors are represented graphically in Figure 18 by the conventional (geometric) vectors $\mathbf{v}_1 = \mathbf{i} - \mathbf{j}$ and $\mathbf{v}_2 = \mathbf{i} + \mathbf{j}$.

As you can see, these two vectors are at right angles. Thus their scalar product is zero: $\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{i} \cdot \mathbf{i} - \mathbf{j} \cdot \mathbf{j} = 0$.

In fact, we do not have to write the eigenvectors in terms of $\mathbf{i}$ and $\mathbf{j}$ to see this, since we can use matrix algebra to define the scalar product. Recall that (in two dimensions) the scalar product of two vectors $\mathbf{v} = [v_1 \quad v_1]^T$ and $\mathbf{w} = [w_1 \quad w_2]^T$ in component form is $\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2$. However, using matrix algebra we see that the scalar product is simply

$$\mathbf{v}^T \mathbf{w} = [v_1 \quad v_2] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = v_1 w_1 + v_2 w_2,$$

and the vectors are perpendicular if $\mathbf{v}^T \mathbf{w} = 0$.

So for the two eigenvectors $\mathbf{v}_1 = [1 \quad -1]^T$ and $\mathbf{v}_2 = [1 \quad 1]^T$, the scalar product is

$$\mathbf{v}_1^T \mathbf{v}_2 = [1 \quad -1] \begin{bmatrix} 1 \\ 1 \end{bmatrix} = (1)(1) + (-1)(1) = 0,$$

and they are indeed perpendicular.

Clearly, this definition of the scalar product works for row vectors of *any* dimension. But in dimensions higher than 3, we tend to call the scalar product the **inner product**, and if the inner product of two vectors vanishes, we tend to say that they are **orthogonal** rather than perpendicular. (This was mentioned in Subsection 2.1 of Unit 4.)

Inner products and orthogonality

Given any two column matrices $\mathbf{v}$ and $\mathbf{w}$, their inner product is given by $\mathbf{v}^T\mathbf{w}$.

If $\mathbf{v}^T\mathbf{w} = 0$, then we say that $\mathbf{v}$ and $\mathbf{w}$ are orthogonal.

Returning to the case of the real symmetric matrix $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, the orthogonality of the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ is not a coincidence. It turns out that *for* any *real symmetric matrix, eigenvectors that correspond to distinct eigenvalues are always orthogonal.* A proof of this is given in Subsection 4.2, but you are not required to know it.

When eigenvalues are not distinct, it turns out that there are extra parameters in the eigenvectors, other than the scalar multiple, and we can always choose values for these parameters to make the eigenvectors orthogonal. This gives us the following powerful statement about eigenvectors of symmetric matrices.

For example, the 2×2 real symmetric matrix $\mathbf{A} = \mathbf{I}$ has two unit eigenvalues, and the eigenvectors are of the form $[x \quad y]^T$ for any x and any y. So we simply choose $x = 1$, $y = 0$ for $\mathbf{v}_1$ and $x = 0$, $y = 1$ for $\mathbf{v}_2$, giving $\mathbf{v}_1 = [1 \quad 0]^T$ and $\mathbf{v}_2 = [0 \quad 1]^T$.

For *any* real symmetric $n \times n$ matrix $\mathbf{A}$, it is always possible to find a set of n real orthogonal eigenvectors of $\mathbf{A}$.

We will use this result in Unit 7 to help us to classify the stationary points of a function of two variables.

Exercise 36

Find the inner product of each of the following pairs of column vectors, and hence determine which, if any, are orthogonal.

(a) $\mathbf{s}_1 = [2 \quad 1]^T$ and $\mathbf{s}_2 = [3 \quad -6]^T$

(b) $\mathbf{t}_1 = [2 \quad 2 \quad 1]^T$ and $\mathbf{t}_2 = [-2 \quad -2 \quad 0]^T$

Exercise 37

Calculate the eigenvectors of $\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$, and show that they are orthogonal.

Let us summarise our results for real symmetric matrices.

Eigenvalues and eigenvectors of a real symmetric matrix

For a real symmetric matrix:

- the eigenvalues are real
- the eigenvectors may be chosen to be real and orthogonal.

Figure 19 Paul Dirac (1902–1984), one of the founders of quantum mechanics

Symmetric matrices – a complex perspective

Complex matrices are used in many applications, and are particularly important in quantum physics, which was developed in the 1920s by many people, including Paul Dirac (Figure 19). There, the closest analogue to a real symmetric matrix is known as a *Hermitian matrix*, which is defined by the requirement that the matrix should equal the complex conjugate of its own transpose (i.e. $(\mathbf{A}^T)^* = \mathbf{A}$, where the star indicates complex conjugation of all the elements). Clearly, if a Hermitian matrix has only real elements, then it is a symmetric matrix. Hermitian matrices also have real eigenvalues. Indeed, the proof (given in Subsection 4.2) that the eigenvalues of a real symmetric matrix are real is directly applicable to the case of Hermitian matrices. Nor does the similarity end there. An $n \times n$ Hermitian matrix also possesses n orthogonal eigenvectors, though they will generally be complex. Hermitian matrices, together with their eigenvalues and eigenvectors, are of fundamental importance in quantum physics, where they describe the dynamics of matter at the atomic and subatomic scale.

4 Further results and proofs regarding real symmetric matrices

The material in this section is optional, and will not be assessed.

Subsection 4.1 extends our analysis of symmetric matrices, showing how one can use the orthogonality of eigenvectors to construct a simple formula for the eigenvector expansion. Subsection 4.2 contains the proofs that the eigenvalues of a real symmetric matrix are real and the eigenvectors are orthogonal.

4.1 Real orthonormal bases

In Subsection 2.4 we introduced the *basis* of a vector space as a generalisation of the idea of Cartesian unit vectors. However, Cartesian unit vectors are especially easy to work with because they are *mutually orthogonal* (implying $\mathbf{i} \cdot \mathbf{j} = \mathbf{j} \cdot \mathbf{k} = \mathbf{i} \cdot \mathbf{k} = 0$), and *normalised* (implying $\mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = 1$). These properties are not shared by the basis of a general vector space.

However, we have now seen that an $n \times n$ real symmetric matrix has n real eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ that may be chosen to be mutually orthogonal in the sense that $\mathbf{v}_i^T \mathbf{v}_j = 0$. Moreover, each of those eigenvectors can be

normalised in a way that is directly analogous to the way in which we form normalised Cartesian unit vectors (by dividing each vector by its own magnitude). In the case of a real eigenvector represented by a column vector $\mathbf{v}_i$, we define normalisation as follows.

Normalisation

Given a real eigenvector $\mathbf{v}_i$, the corresponding **normalised eigenvector** $\widehat{\mathbf{v}}_i$ is given by (see Subsection 1.4 of Unit 4)

$$\widehat{\mathbf{v}}_i = \frac{\mathbf{v}_i}{|\mathbf{v}_i|} = \frac{\mathbf{v}_i}{\sqrt{\mathbf{v}_i^T \mathbf{v}_i}}, \qquad (22)$$

and it satisfies the **normalisation condition** $\widehat{\mathbf{v}}_i^T \widehat{\mathbf{v}}_i = 1$.

Here, the magnitude of $\mathbf{v}$ has been defined as $|\mathbf{v}| = \sqrt{\mathbf{v}_i^T \mathbf{v}_i}$.

Exercise 38

In Exercise 37 you showed that the matrix $\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$ has orthogonal eigenvectors $\mathbf{v}_1 = [2 \quad 1]^T$ and $\mathbf{v}_2 = [1 \quad -2]^T$. Write down the corresponding normalised eigenvectors.

A set of vectors in which each element is normalised as well as being orthogonal to all the other vectors belonging to the set, is said to be **orthonormal**; e.g. $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are orthonormal. We therefore arrive at the following important conclusion regarding the eigenvectors of a real symmetric matrix.

Given any $n \times n$ real symmetric matrix, it is always possible to find a set of n real orthonormal eigenvectors $\widehat{\mathbf{v}}_1, \widehat{\mathbf{v}}_2, \ldots, \widehat{\mathbf{v}}_n$ for which

$$\widehat{\mathbf{v}}_i^T \widehat{\mathbf{v}}_j = 1 \quad \text{if } i = j \qquad \text{and} \qquad \widehat{\mathbf{v}}_i^T \widehat{\mathbf{v}}_j = 0 \quad \text{if } i \neq j.$$

Such a set is linearly independent, and therefore forms an **orthonormal basis** for the n-dimensional vector space.

It follows from this that for any n-dimensional column vector $\mathbf{v}$, we have the eigenvector expansion

$$\mathbf{v} = \alpha_1 \widehat{\mathbf{v}}_1 + \alpha_2 \widehat{\mathbf{v}}_2 + \cdots + \alpha_n \widehat{\mathbf{v}}_n. \qquad (23)$$

We have seen eigenvector expansions before, but in those earlier cases the eigenvectors concerned were just linearly independent, not orthonormal. So working out the scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$ was not easy and was largely avoided except for $n = 2$. However, orthonormality makes things much simpler. If we matrix multiply each side of equation (23) on the left by $\widehat{\mathbf{v}}_i^T$, we get

$$\widehat{\mathbf{v}}_i^T \mathbf{v} = \alpha_1 \widehat{\mathbf{v}}_i^T \widehat{\mathbf{v}}_1 + \alpha_2 \widehat{\mathbf{v}}_i^T \widehat{\mathbf{v}}_2 + \cdots + \alpha_n \widehat{\mathbf{v}}_i^T \widehat{\mathbf{v}}_n.$$

Then, using the orthogonality of the normalised eigenvectors, all of the inner products on the right-hand side (i.e. all the matrix products of the form $\widehat{\mathbf{v}}_i^T\widehat{\mathbf{v}}_j$) must vanish, apart from the one for which $j = i$. Thus

$$\widehat{\mathbf{v}}_i^T\mathbf{v} = \alpha_i\widehat{\mathbf{v}}_i^T\widehat{\mathbf{v}}_i.$$

We can now use the fact that each of the orthonormal eigenvectors has magnitude 1 (i.e. $\widehat{\mathbf{v}}_i^T\widehat{\mathbf{v}}_i = 1$) to write

$$\widehat{\mathbf{v}}_i^T\mathbf{v} = \alpha_i.$$

This equation gives all the coefficients α_i for $i = 1, 2, \ldots, n$. Hence equation (23) can be rewritten as follows.

Real orthonormal eigenvector expansion

$$\mathbf{v} = (\widehat{\mathbf{v}}_1^T\mathbf{v})\widehat{\mathbf{v}}_1 + (\widehat{\mathbf{v}}_2^T\mathbf{v})\widehat{\mathbf{v}}_2 + \cdots + (\widehat{\mathbf{v}}_n^T\mathbf{v})\widehat{\mathbf{v}}_n. \tag{24}$$

This is a very useful expression that enables us to represent a vector as an eigenvector expansion of a real symmetric matrix.

Of course, all of this is expressed in the rather general language of linear algebra, but it should be clear that the real orthonormal eigenvectors $\widehat{\mathbf{v}}_i$ are nothing more than generalisations of the familiar Cartesian unit vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$, which are also orthonormal. And equation (24) is simply a generalisation of $\mathbf{v} = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k}$, where $v_x = \mathbf{i}\cdot\mathbf{v}$, $v_y = \mathbf{j}\cdot\mathbf{v}$ and $v_z = \mathbf{k}\cdot\mathbf{v}$.

Here is a numerical application of equation (24).

Example 15

The real symmetric matrix $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ has eigenvectors $\mathbf{v}_1 = [1 \quad -1]^T$ and $\mathbf{v}_2 = [1 \quad 1]^T$, corresponding to the eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 3$.

(a) Find the normalised orthogonal eigenvectors $\widehat{\mathbf{v}}_1$ and $\widehat{\mathbf{v}}_2$.

(b) Express the vector $\mathbf{v} = [3 \quad -2]^T$ as an eigenvector expansion using the orthonormal basis consisting of $\widehat{\mathbf{v}}_1$ and $\widehat{\mathbf{v}}_2$.

Solution

(a) Since $\mathbf{v}_1 = [1 \quad -1]^T$ and $\mathbf{v}_2 = [1 \quad 1]^T$ are already orthogonal, we only need to normalise them:

$$\widehat{\mathbf{v}}_1 = \frac{\mathbf{v}_1}{\sqrt{\mathbf{v}_1^T\mathbf{v}_1}} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \widehat{\mathbf{v}}_2 = \frac{\mathbf{v}_2}{\sqrt{\mathbf{v}_2^T\mathbf{v}_2}} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

(b) From equation (24), $\mathbf{v} = (\widehat{\mathbf{v}}_1^T\mathbf{v})\widehat{\mathbf{v}}_1 + (\widehat{\mathbf{v}}_2^T\mathbf{v})\widehat{\mathbf{v}}_2$. Calculating the coefficients, we have

$$\widehat{\mathbf{v}}_1^T\mathbf{v} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\end{bmatrix}\begin{bmatrix}3\\-2\end{bmatrix} = \frac{5}{\sqrt{2}} \quad \text{and} \quad \widehat{\mathbf{v}}_2^T\mathbf{v} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}3\\-2\end{bmatrix} = \frac{1}{\sqrt{2}}.$$

Hence the eigenvector expansion is

$$\mathbf{v} = \frac{5}{\sqrt{2}}\widehat{\mathbf{v}}_1 + \frac{1}{\sqrt{2}}\widehat{\mathbf{v}}_2.$$

We can easily check this by substituting the values for $\mathbf{v}_1$ and $\mathbf{v}_2$:

$$\mathbf{v} = \frac{5}{2}\begin{bmatrix}1\\-1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}6\\-4\end{bmatrix} = \begin{bmatrix}3\\-2\end{bmatrix},$$

as required.

Exercise 39

In Exercise 38 you showed that $\widehat{\mathbf{v}}_1 = \frac{1}{\sqrt{5}}[2 \quad 1]^T$ and $\widehat{\mathbf{v}}_2 = \frac{1}{\sqrt{5}}[1 \quad -2]^T$ are orthonormal eigenvectors of the matrix $\mathbf{A} = \begin{bmatrix}5 & 2\\2 & 2\end{bmatrix}$. Express the vector $\mathbf{v} = [1 \quad 1]^T$ in terms of this basis.

4.2 Proofs concerning real symmetric matrices

Proof that for real symmetric matrices the eigenvalues are real

Throughout the proof, stars are used to indicate complex conjugation. So given a complex quantity $z = a + ib$, its complex conjugate is $z^* = (a + ib)^* = (a - ib)$. Note that $z^*z = a^2 + b^2 = |z|^2$. The use of a star to indicate complex conjugation is a very common alternative to the overline that is used elsewhere in the module.

Let $\mathbf{A}$ be a real symmetric matrix, and let λ be one of its eigenvalues that corresponds to an eigenvector $\mathbf{v}$. In such a situation,

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}. \tag{25}$$

We know that the elements of $\mathbf{A}$ are real, but at this stage we cannot be sure that either λ or $\mathbf{v}$ is necessarily real. Taking the transpose of both sides of equation (25) gives

The rule for taking the transpose of a product of matrices was discussed in Unit 4.

$$\mathbf{v}^T\mathbf{A}^T = \lambda\mathbf{v}^T.$$

Taking the complex conjugate of both sides then gives

$$(\mathbf{v}^T)^*(\mathbf{A}^T)^* = \lambda^*(\mathbf{v}^T)^*.$$

But $\mathbf{A}$ is both real and symmetric, so $(\mathbf{A}^T)^* = \mathbf{A}$, giving

$$(\mathbf{v}^T)^*\mathbf{A} = \lambda^*(\mathbf{v}^T)^*.$$

Matrix multiplying each side of this equation on the right by $\mathbf{v}$ then gives

$$(\mathbf{v}^T)^*\mathbf{A}\mathbf{v} = \lambda^*(\mathbf{v}^T)^*\mathbf{v}.$$

On the other hand, matrix multiplying each side of equation (25) on the left by $(\mathbf{v}^T)^*$ gives

$$(\mathbf{v}^T)^*\mathbf{A}\mathbf{v} = \lambda(\mathbf{v}^T)^*\mathbf{v}.$$

Subtracting these last two equations gives

$$0 = (\lambda - \lambda^*)(\mathbf{v}^T)^*\mathbf{v}. \tag{26}$$

However, if $\mathbf{v} = [v_1 \quad v_2 \quad \ldots \quad v_n]^T$, then

$$(\mathbf{v}^T)^*\mathbf{v} = v_1^* v_1 + v_2^* v_2 + \cdots + v_n^* v_n = |v_1|^2 + |v_2|^2 + \cdots + |v_n|^2,$$

which must be a positive quantity and non-zero because eigenvectors are always non-zero. Since $(\mathbf{v}^T)^*\mathbf{v}$ is positive, it follows from equation (26) that

$$0 = \lambda - \lambda^*,$$

from which we see that $\lambda = \lambda^*$, so λ must be real.

Proof that for real symmetric matrices the eigenvectors corresponding to distinct eigenvalues are orthogonal

Given that $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ and $\mathbf{A}\mathbf{v}_j = \lambda_j \mathbf{v}_j$, taking the transpose of the equation involving $\mathbf{v}_i$ gives

$$\mathbf{v}_i^T \mathbf{A}^T = \lambda_i \mathbf{v}_i^T,$$

and matrix multiplying each side of this on the right by $\mathbf{v}_j$ gives

$$\mathbf{v}_i^T \mathbf{A}^T \mathbf{v}_j = \lambda_i \mathbf{v}_i^T \mathbf{v}_j.$$

Since $\mathbf{A}$ is symmetric, so that $\mathbf{A}^T = \mathbf{A}$, this may be rewritten as

$$\mathbf{v}_i^T \mathbf{A} \mathbf{v}_j = \lambda_i \mathbf{v}_i^T \mathbf{v}_j.$$

On the other hand, starting from the equation $\mathbf{A}\mathbf{v}_j = \lambda_j \mathbf{v}_j$, and matrix multiplying each side on the left by $\mathbf{v}_i^T$, we see that

$$\mathbf{v}_i^T \mathbf{A} \mathbf{v}_j = \lambda_j \mathbf{v}_i^T \mathbf{v}_j.$$

Subtracting the last two equations, we find

$$0 = (\lambda_i - \lambda_j)\mathbf{v}_i^T \mathbf{v}_j. \tag{27}$$

However, λ_i and λ_j are distinct, so $\lambda_i - \lambda_j \neq 0$. It therefore follows from equation (27) that $\mathbf{v}_i^T \mathbf{v}_j = 0$, which is just the condition for the real eigenvectors $\mathbf{v}_i$ and $\mathbf{v}_j$ to be orthogonal.

Learning outcomes

After studying this unit, you should be able to do the following.

- Solve 2×2 and 3×3 systems of linear equations by the Gaussian elimination method, using row interchanges where necessary.
- Determine whether such a system of equations has no solution, a unique solution, or an infinity of solutions.
- Explain the meaning of the terms eigenvector, eigenvalue, linearly independent and basis.
- Understand how repeated application of a matrix to a vector becomes proportional to the eigenvector with largest modulus eigenvalue.
- Use the characteristic equation and eigenvector equation to calculate the eigenvalues and eigenvectors of a 2×2 matrix.
- Calculate the eigenvalues and eigenvectors of a 3×3 matrix, where one of the eigenvalues is obvious or given.
- Appreciate that an $n \times n$ matrix with n distinct eigenvalues gives rise to n linearly independent eigenvectors that can be used as a basis.
- Know how to calculate an eigenvector expansion of a given vector.
- Appreciate that the eigenvalues of a matrix may be real or complex, and may be distinct or repeated.
- Appreciate that the complex eigenvalues and corresponding eigenvectors of real matrices occur in complex conjugate pairs.
- Write down the eigenvalues of a triangular matrix.
- Recall that the sum of the eigenvalues of a matrix $\mathbf{A}$ is $\operatorname{tr} \mathbf{A}$, and that the product is $\det \mathbf{A}$, and use these properties to help to check and determine eigenvalues.
- Know that a real symmetric matrix has real eigenvalues and that the eigenvectors can be chosen to be real and orthogonal.

Solutions to exercises

Solution to Exercise 1

Yes, the Cartesian equation of a plane is linear in x, y and z. This can be seen by comparing the Cartesian equation with the general linear equation for $n = 3$ and making the identifications $a_1 = a$, $x_1 = x$, $a_2 = b$, $x_2 = y$, $a_3 = c$, $x_3 = z$ and $d = 0$.

Solution to Exercise 2

(a) The required matrix form is

$$\begin{bmatrix} 1 & 1 & -1 \\ 5 & 2 & 2 \\ 4 & -2 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 20 \\ 15 \end{bmatrix}.$$

(b) The equations are just a messy rearrangement of those in part (a), so after further rearrangement to return them to the form given in part (a), we get the same answer as before.

(c) Remembering to insert zero coefficients in place of missing terms, the required matrix form is

$$\begin{bmatrix} 2 & 3 & -4 \\ 2 & 0 & 3 \\ 0 & 6 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}.$$

Solution to Exercise 3

The augmented matrix is

$$\left[\begin{array}{rrr|r} 1 & 1 & -1 & 2 \\ 5 & 2 & 2 & 20 \\ 4 & -2 & -3 & 15 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array}.$$

First, we reduce the elements below the leading diagonal in column 1 to zero:

$$\begin{array}{r} \\ \mathbf{R}_2 - 5\mathbf{R}_1 \\ \mathbf{R}_3 - 4\mathbf{R}_1 \end{array} \left[\begin{array}{rrr|r} 1 & 1 & -1 & 2 \\ 0 & -3 & 7 & 10 \\ 0 & -6 & 1 & 7 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array}.$$

Then we reduce the element below the leading diagonal in column 2 to zero:

$$\begin{array}{r} \\ \\ \mathbf{R}_{3a} - 2\mathbf{R}_{2a} \end{array} \left[\begin{array}{rrr|r} 1 & 1 & -1 & 2 \\ 0 & -3 & 7 & 10 \\ 0 & 0 & -13 & -13 \end{array}\right].$$

The equations represented by the new matrix are

$$\begin{aligned} x_1 + x_2 - x_3 &= 2, \\ -3x_2 + 7x_3 &= 10, \\ -13x_3 &= -13, \end{aligned}$$

which give $x_3 = 1$, $x_2 = \frac{1}{3}(7x_3 - 10) = -1$ and $x_1 = 2 - x_2 + x_3 = 4$.

We verify the solution as follows:

$$\begin{bmatrix} 1 & 1 & -1 \\ 5 & 2 & 2 \\ 4 & -2 & -3 \end{bmatrix} \begin{bmatrix} 4 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4-1-1 \\ 20-2+2 \\ 16+2-3 \end{bmatrix} = \begin{bmatrix} 2 \\ 20 \\ 15 \end{bmatrix}.$$

Solution to Exercise 4

Let the salaries (in thousands of rurs) of managers, software engineers and clerks at a bank be x_i, where $i = 1, 2, 3$, respectively. The problem is then specified by the simultaneous equations

$$\begin{aligned} 3x_1 + 2x_2 + 24x_3 &= 137, \\ x_1 + x_2 + 26x_3 &= 137, \\ 3x_2 + 25x_3 &= 137. \end{aligned}$$

This problem is unaffected by exchanging the order of the first two equations, to obtain the augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 1 & 26 & 137 \\ 3 & 2 & 24 & 137 \\ 0 & 3 & 25 & 137 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array},$$

from which we obtain

$$\begin{array}{r} \\ \mathbf{R}_2 - 3\mathbf{R}_1 \\ \\ \end{array} \left[\begin{array}{ccc|c} 1 & 1 & 26 & 137 \\ 0 & -1 & -54 & -274 \\ 0 & 3 & 25 & 137 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2\mathrm{a}} \\ \mathbf{R}_{3\mathrm{a}} \end{array}$$

without introducing fractional elements. We complete the elimination stage as follows:

$$\begin{array}{r} \\ \\ \mathbf{R}_{3\mathrm{a}} + 3\mathbf{R}_{2\mathrm{a}} \end{array} \left[\begin{array}{ccc|c} 1 & 1 & 26 & 137 \\ 0 & -1 & -54 & -274 \\ 0 & 0 & -137 & -685 \end{array}\right].$$

From this we see that $x_3 = 5$ for clerks, and back substitution gives $x_2 = 274 - 54x_3 = 4$ for software engineers, and $x_1 = 137 - x_2 - 26x_3 = 3$ for managers. Thus the numbers required are given by the vector $\mathbf{x} = [3 \quad 4 \quad 5]^T$, which can be confirmed as a solution of the original matrix equation.

So in Ruritania, clerks receive 5000 rurs per month, software engineers receive 4000 rurs per month, and managers receive 3000 rurs per month.

Solution to Exercise 5

Starting with

$$\left[\begin{array}{ccc|c} 0 & 0 & 1 & 2 \\ 2 & 1 & 0 & 5 \\ 1 & 2 & 0 & 7 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array},$$

we eliminate x_1 from the second row, to obtain

$$\begin{array}{r} \\ \mathbf{R}_2 - 2\mathbf{R}_3 \\ \\ \end{array} \left[\begin{array}{ccc|c} 0 & 0 & 1 & 2 \\ 0 & -3 & 0 & -9 \\ 1 & 2 & 0 & 7 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_3 \end{array}.$$

Then $\mathbf{R}_1$ gives $x_3 = 2$, $\mathbf{R}_{2a}$ gives $x_2 = 3$, and $\mathbf{R}_3$ gives $x_1 = 7 - 2x_2 = 1$. Hence the solution is $\mathbf{x} = [1 \quad 3 \quad 2]^T$. (As usual, it should be confirmed that the original matrix equation is satisfied by this solution.)

Solution to Exercise 6

Starting with

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 10 \\ 1 & 1 & 2 & 1 & 12 \\ 1 & 3 & 1 & 1 & 16 \\ 4 & 1 & 1 & 1 & 22 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \\ \mathbf{R}_4 \end{array},$$

we eliminate x_1 from the last three rows, to obtain

$$\begin{array}{r} \\ \mathbf{R}_2 - \mathbf{R}_1 \\ \mathbf{R}_3 - \mathbf{R}_1 \\ \mathbf{R}_4 - 4\mathbf{R}_1 \end{array} \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 10 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 2 & 0 & 0 & 6 \\ 0 & -3 & -3 & -3 & -18 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \\ \mathbf{R}_{4a} \end{array}.$$

We could interchange rows to obtain an upper triangular augmented matrix, but it should also be clear that $\mathbf{R}_{2a}$ gives $x_3 = 2$, $\mathbf{R}_{3a}$ gives $x_2 = 3$, $\mathbf{R}_{4a}$ gives $x_4 = 6 - x_2 - x_3 = 1$, and $\mathbf{R}_1$ gives $x_1 = 10 - x_2 - x_3 - x_4 = 4$. (As usual, it should be confirmed that the original matrix equation is satisfied by the solution $\mathbf{x} = [4 \quad 3 \quad 2 \quad 1]^T$.)

Solution to Exercise 7

(a) From the augmented matrix

$$\left[\begin{array}{ccc|c} 1 & -2 & 5 & 7 \\ 1 & 3 & -4 & 20 \\ 1 & 18 & -31 & 40 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array},$$

we obtain

$$\begin{array}{r} \\ \mathbf{R}_2 - \mathbf{R}_1 \\ \mathbf{R}_3 - \mathbf{R}_1 \end{array} \left[\begin{array}{ccc|c} 1 & -2 & 5 & 7 \\ 0 & 5 & -9 & 13 \\ 0 & 20 & -36 & 33 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array}$$

and

$$\begin{array}{r} \\ \\ \mathbf{R}_{3a} - 4\mathbf{R}_{2a} \end{array} \left[\begin{array}{ccc|c} 1 & -2 & 5 & 7 \\ 0 & 5 & -9 & 13 \\ 0 & 0 & 0 & -19 \end{array}\right].$$

The third equation reads $0 = -19$, which is impossible. So the equations are inconsistent and there is no solution.

(b) From the augmented matrix

$$\begin{array}{c} \left[\begin{array}{ccc|c} 1 & -2 & 5 & 6 \\ 1 & 3 & -4 & 7 \\ 2 & 6 & -12 & 12 \end{array}\right] \end{array} \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array},$$

we obtain

$$\begin{array}{r} \\ \mathbf{R}_2 - \mathbf{R}_1 \\ \mathbf{R}_3 - 2\mathbf{R}_1 \end{array} \left[\begin{array}{ccc|c} 1 & -2 & 5 & 6 \\ 0 & 5 & -9 & 1 \\ 0 & 10 & -22 & 0 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array}$$

and

$$\begin{array}{r} \\ \\ \mathbf{R}_{3a} - 2\mathbf{R}_{2a} \end{array} \left[\begin{array}{ccc|c} 1 & -2 & 5 & 6 \\ 0 & 5 & -9 & 1 \\ 0 & 0 & -4 & -2 \end{array}\right],$$

then back substitution gives a unique solution.

(c) From the augmented matrix

$$\left[\begin{array}{ccc|c} 1 & -4 & 1 & 14 \\ 5 & -1 & -1 & 2 \\ 6 & 14 & -6 & -52 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array},$$

we obtain

$$\begin{array}{r} \\ \mathbf{R}_2 - 5\mathbf{R}_1 \\ \mathbf{R}_3 - 6\mathbf{R}_1 \end{array} \left[\begin{array}{ccc|c} 1 & -4 & 1 & 14 \\ 0 & 19 & -6 & -68 \\ 0 & 38 & -12 & -136 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array}$$

and

$$\begin{array}{r} \\ \\ \mathbf{R}_{3a} - 2\mathbf{R}_{2a} \end{array} \left[\begin{array}{ccc|c} 1 & -4 & 1 & 14 \\ 0 & 19 & -6 & -68 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The third equation reads $0 = 0$, which is true but does not provide any limitation on the possible value of x_3. So there is an infinity of solutions. In this case we may assign x_3 any real non-zero value that we choose, say $x_3 = p$. Then from the second equation we get $x_2 = \frac{1}{19}(-68 + 6p)$, and from the first equation we get $x_1 = 14 + 4x_2 - x_3 = \frac{1}{19}(-6 + 5p)$.

The existence of an infinity of solutions is indicated by the infinity of possible choices for p.

Solution to Exercise 8

Matrix multiplication shows that the column vectors given in (b) and (d) are both eigenvectors (corresponding to the eigenvalue 1), but those given in (a) and (c) are not. The results of the matrix multiplications in the four cases are as follows.

(a) $\begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} 1000 \\ 1000 \end{bmatrix} = \begin{bmatrix} 1000 \\ 900 \end{bmatrix}$

(b) $\begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} 120 \\ 60 \end{bmatrix} = \begin{bmatrix} 120 \\ 60 \end{bmatrix}$

(c) $\begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} 500 \\ 300 \end{bmatrix} = \begin{bmatrix} 510 \\ 290 \end{bmatrix}$

(d) $\begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} 20 \\ 10 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \end{bmatrix}$

Solution to Exercise 9

The scaled vector $k\mathbf{w}$ is represented by the column vector $\mathbf{v} = k\mathbf{w} = [k \quad k]^T$, so

$$\mathbf{A}\mathbf{v} = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} k \\ k \end{bmatrix} = \begin{bmatrix} 5k \\ 5k \end{bmatrix} = 5\mathbf{v}.$$

So $\mathbf{A}$ does indeed map $k\mathbf{w}$ to $5k\mathbf{w}$. This shows that any scaled vector $\mathbf{v} = k\mathbf{w}$ is an eigenvector of $\mathbf{A}$ corresponding to the eigenvalue 5.

Solution to Exercise 10

(a) $$\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 12 \\ 8 \end{bmatrix} = 4 \begin{bmatrix} 3 \\ 2 \end{bmatrix},$$

so $[3 \quad 2]^T$ is an eigenvector with eigenvalue 4.

(b) $$\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} = (-1) \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

so $[1 \quad -1]^T$ is an eigenvector with eigenvalue -1.

(c) $$\begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 12 \end{bmatrix} = 2 \begin{bmatrix} 0 \\ 6 \end{bmatrix},$$

so $[0 \quad 6]^T$ is an eigenvector with eigenvalue 2.

(d) $$\begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ -2 \end{bmatrix},$$

so $[1 \quad -2]^T$ is an eigenvector with eigenvalue 0.

Solution to Exercise 11

$$\begin{bmatrix} 3 & -2 \\ 4 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1-i \end{bmatrix} = \begin{bmatrix} 1+2i \\ 3+i \end{bmatrix} = (1+2i) \begin{bmatrix} 1 \\ 1-i \end{bmatrix}.$$

So $1+2i$ is an eigenvalue and corresponds to eigenvector $\mathbf{v}_1$.

$$\begin{bmatrix} 3 & -2 \\ 4 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1+i \end{bmatrix} = \begin{bmatrix} 1-2i \\ 3-i \end{bmatrix} = (1-2i) \begin{bmatrix} 1 \\ 1+i \end{bmatrix}.$$

So $1-2i$ is an eigenvalue and corresponds to eigenvector $\mathbf{v}_2$.

Solution to Exercise 12

The eigenvectors act along the line of reflection $y = x$ and perpendicular to it, so they are the scalar multiples of $[1 \quad 1]^T$ and $[1 \quad -1]^T$. The vector $[1 \quad 1]^T$ is scaled by a factor of 1 by the transformation, while for $[1 \quad -1]^T$ the scale factor is -1; these scale factors are the corresponding eigenvalues.

We may check our conclusion by evaluating

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

so $[1 \quad 1]^T$ corresponds to the eigenvalue 1, and

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

so $[1 \quad -1]^T$ corresponds to the eigenvalue -1.

Solution to Exercise 13

Explicit matrix multiplication shows that

$$\begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

so $[2 \quad 1]^T$ is an eigenvector with eigenvalue 1.

Similarly,

$$\begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0.7 \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

so $[1 \quad -1]^T$ is an eigenvector with eigenvalue 0.7.

Solution to Exercise 14

If $\alpha_1 \neq 0$, then rearranging equation (6) gives

$$\mathbf{v}_1 = -\frac{\alpha_2}{\alpha_1}\mathbf{v}_2 - \frac{\alpha_3}{\alpha_1}\mathbf{v}_3 - \cdots - \frac{\alpha_n}{\alpha_1}\mathbf{v}_n.$$

Solution to Exercise 15

(a) $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are certainly linear independent, since they are not coplanar.

(b) These vectors are linearly dependent. The first vector is -1 times the second vector, so those two vectors are antiparallel, and the three vectors are coplanar. Also, the system fails the test for linear independence since

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \alpha_3\mathbf{v}_3 = \mathbf{0}$$

is satisfied, for example, by the values $\alpha_1 = 1$, $\alpha_2 = -1$, $\alpha_3 = 0$, which are not all zero.

Solution to Exercise 16

These vectors are linearly dependent, as there cannot be more than three linearly independent vectors in a three-dimensional space.

In fact, the vectors are related as follows: $\mathbf{v}_4 = -\frac{2}{3}\mathbf{v}_1 + \frac{2}{3}\mathbf{v}_2 + \frac{2}{3}\mathbf{v}_3$.

Solution to Exercise 17

Setting $\mathbf{v} = \alpha\mathbf{v}_1 + \beta\mathbf{v}_2$, we have

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} -2 \\ 1 \end{bmatrix},$$

from which we get the simultaneous linear equations

$$\begin{aligned} 1 &= \alpha - 2\beta, \\ 3 &= \alpha + \ \ \beta. \end{aligned}$$

Solving these, we obtain $\alpha = 7/3$ and $\beta = 2/3$.

Solution to Exercise 18

From the definition of linear independence (see equation (6)), we need to show that the only solution of $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \alpha_3\mathbf{v}_3 = \mathbf{0}$ is $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

We have

$$\alpha_1 \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + \alpha_3 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

This gives a system of three equations for $\alpha_1, \alpha_2, \alpha_3$, which we put in augmented matrix form:

$$\left[\begin{array}{ccc|c} -1 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array}.$$

We now solve these equations by Gaussian elimination:

$$\begin{array}{r} \\ \mathbf{R}_2 + \mathbf{R}_1 \\ \mathbf{R}_3 + \mathbf{R}_1 \end{array} \left[\begin{array}{ccc|c} -1 & 1 & 0 & 0 \\ 0 & 3 & 2 & 0 \\ 0 & 2 & 1 & 0 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array},$$

$$\begin{array}{l} \\ \\ \mathbf{R}_{3a} - \frac{2}{3}\mathbf{R}_{2a} \end{array} \left[\begin{array}{ccc|c} -1 & 1 & 0 & 0 \\ 0 & 3 & 2 & 0 \\ 0 & 0 & -\frac{1}{3} & 0 \end{array}\right].$$

Back substitution gives $\alpha_1 = \alpha_2 = \alpha_3 = 0$, hence the eigenvectors are linearly independent.

Solution to Exercise 19

(a) The eigenvectors correspond to different eigenvalues, so are distinct. This is also obvious because they are not collinear. They therefore form a basis for two-dimensional vectors, so we can write

$$\mathbf{r}_0 = \begin{bmatrix} -2 \\ 4 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 - 2c_2 \\ c_1 + c_2 \end{bmatrix}.$$

Equating corresponding elements of the column vectors on the right and the left shows that

$$c_1 - 2c_2 = -2 \quad \text{and} \quad c_1 + c_2 = 4.$$

Solving this simple system of linear equations, we see that $c_1 = 2$ and $c_2 = 2$. Thus

$$\mathbf{r}_0 = 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 2\begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

(b) Using the eigenvector expansion gives

$$\mathbf{A}\begin{bmatrix} -2 \\ 4 \end{bmatrix} = \mathbf{A}\left(2\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 2\begin{bmatrix} -2 \\ 1 \end{bmatrix}\right) = 10\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 4\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 14 \end{bmatrix}.$$

Solution to Exercise 20

From the solution to Exercise 19, we have

$$\mathbf{r}_0 = \begin{bmatrix} -2 \\ 4 \end{bmatrix} = 2\mathbf{v}_1 + 2\mathbf{v}_2,$$

where $\mathbf{v}_1 = [1 \quad 1]^T$ and $\mathbf{v}_2 = [-2 \quad 1]^T$ are the eigenvectors of $\mathbf{A}$ that correspond to the real eigenvalues $\lambda_1 = 5$ and $\lambda_2 = 2$. Hence

$$\begin{aligned} \mathbf{r}_8 = \mathbf{A}^8\mathbf{r}_0 &= 2\mathbf{A}^8\mathbf{v}_1 + 2\mathbf{A}^8\mathbf{v}_2 \\ &= 2\lambda_1^8\mathbf{v}_1 + 2\lambda_2^8\mathbf{v}_2 \\ &= 2(5^8)\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 2(2^8)\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 780\,226 \\ 781\,762 \end{bmatrix}. \end{aligned}$$

Thus to two significant figures,

$$\mathbf{r}_8 = \begin{bmatrix} 78 \times 10^4 \\ 78 \times 10^4 \end{bmatrix} \simeq 2(5^8)\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Similarly,

$$\mathbf{r}_9 = \begin{bmatrix} 39 \times 10^5 \\ 39 \times 10^5 \end{bmatrix} \simeq 2(5^9)\begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{r}_{10} = \begin{bmatrix} 20 \times 10^6 \\ 20 \times 10^6 \end{bmatrix} \simeq 2(5^{10})\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The significance of these results is that, working to two significant figures, the contribution of the eigenvector corresponding to the smaller of the two eigenvalues is negligible in the eighth, ninth and tenth iterations. So to two significant figures, $\mathbf{A}^k\mathbf{x} \simeq 2\lambda_1^k\mathbf{v}_1$ for $k \geq 8$.

Solution to Exercise 21

With $\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$ we obtain

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} 3-\lambda & 2 \\ 1 & 4-\lambda \end{vmatrix} = (3-\lambda)(4-\lambda) - 2,$$

so the characteristic equation may be written as $\lambda^2 - 7\lambda + 10 = 0$, the roots of which, i.e. the eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 5$, may be found using the formula or factorising.

Solution to Exercise 22

(a) The characteristic equation of $\mathbf{G}$ gives

$$\begin{vmatrix} 3-\lambda & 2 \\ 1 & 2-\lambda \end{vmatrix} = (3-\lambda)(2-\lambda) - 2 = 0.$$

This becomes $\lambda^2 - 5\lambda + 4 = 0$. Solving this using the standard formula, we get

$$\lambda = \frac{5 \pm \sqrt{25-16}}{2}.$$

So the eigenvalues of $\mathbf{G}$ are 1 and 4.

(b) The characteristic equation of $\mathbf{H}$ gives

$$\begin{vmatrix} 2-\lambda & 2 \\ -1 & 2-\lambda \end{vmatrix} = (2-\lambda)(2-\lambda) + 2 = 0.$$

This becomes $\lambda^2 - 4\lambda + 6 = 0$. Solving this using the standard formula, we get

$$\lambda = \frac{4 \pm \sqrt{16-24}}{2} = \frac{4 \pm \sqrt{-8}}{2}.$$

So the eigenvalues of $\mathbf{H}$ are $2 + i\sqrt{2}$ and $2 - i\sqrt{2}$.

Solution to Exercise 23

(a) The characteristic equation of $\mathbf{R}$ gives

$$\lambda^2 - (\cos\theta + \cos\theta)\lambda + (\cos^2\theta + \sin^2\theta) = 0.$$

Using the fact that $\cos^2\theta + \sin^2\theta = 1$, we get

$$\lambda^2 - 2\lambda\cos\theta + 1 = 0.$$

Solving this using the standard formula, we get

$$\begin{aligned}\lambda &= \frac{2\cos\theta \pm \sqrt{4\cos^2\theta - 4}}{2}\\ &= \cos\theta \pm \sqrt{\cos^2\theta - 1}\\ &= \cos\theta \pm \sqrt{-\sin^2\theta} = \cos\theta \pm i\sin\theta = e^{\pm i\theta}.\end{aligned}$$

Note that we could have obtained this solution much more quickly by working directly from the characteristic equation:

$$\begin{vmatrix}\cos\theta - \lambda & -\sin\theta\\ \sin\theta & \cos\theta - \lambda\end{vmatrix} = (\cos\theta - \lambda)^2 + \sin^2\theta = 0.$$

This gives $(\cos\theta - \lambda)^2 = -\sin^2\theta$, so taking the square root of both sides gives $(\cos\theta - \lambda) = \pm i\sin\theta$, hence $\lambda = \cos\theta \pm i\sin\theta$. Using Euler's theorem, this can also be expressed as $\lambda = e^{\pm i\theta}$.

(b) The characteristic equation of $\mathbf{M}$ is $(l - \lambda)(k - \lambda) = 0$, hence the eigenvalues are l and k.

Solution to Exercise 24

The characteristic equation of $\mathbf{S}$ is

$$\begin{vmatrix}1-\lambda & s\\ 0 & 1-\lambda\end{vmatrix} = (1-\lambda)^2 = 0,$$

hence we have a pair of repeated eigenvalues $\lambda = 1$.

Solution to Exercise 25

$$\operatorname{tr}\mathbf{A} = 11 + 15 = 26 \quad \text{and} \quad \det\mathbf{A} = 11\times 15 - 14\times 12 = -3.$$

To the required level of accuracy,

$$\lambda_1 + \lambda_2 = 26.00 = \operatorname{tr}\mathbf{A} \quad \text{and} \quad \lambda_1\lambda_2 = -3.00 = \det\mathbf{A}.$$

The given values therefore satisfy the trace and determinant checks for eigenvalues of $\mathbf{A}$.

Solution to Exercise 26

(a) In this case the characteristic equation of $\mathbf{A}$ is given by

$$\det(\mathbf{A}-\lambda\mathbf{I}) = \begin{vmatrix} 1-\lambda & -2 & 4 \\ -2 & 7-\lambda & -10 \\ -1 & 4 & -6-\lambda \end{vmatrix} = 0.$$

Using Laplace's rule to expand the determinant in terms of the elements of the top row gives

$$(1-\lambda)[(7-\lambda)(-6-\lambda)+40] - (-2)[-2(-6-\lambda)-10] \\ + 4[-8+(7-\lambda)] = 0.$$

This gives

$$\lambda^3 - 2\lambda^2 - \lambda + 2 = 0. \tag{28}$$

Given that one of the factors is 2, we factorise this by writing

$$(\lambda-2)(a\lambda^2+b\lambda+c) = 0,$$

for some constants a, b, c. Expanding this, we get

$$a\lambda^3 + (b-2)\lambda^2 + (c-2b)\lambda - 2c = 0.$$

Comparing with equation (28), we see that $a = 1$, $b = 0$ and $c = -1$. So equation (28) can be factorised to give

$$(\lambda-2)(\lambda^2-1) = (\lambda-2)(\lambda-1)(\lambda+1) = 0,$$

hence the eigenvalues are -1, 1 and 2.

(b) In this case the characteristic equation of $\mathbf{A}$ is given by

$$\det(\mathbf{A}-\lambda\mathbf{I}) = \begin{vmatrix} 4-\lambda & 7 & 6 \\ 6 & 5-\lambda & 6 \\ -8 & -10 & -10-\lambda \end{vmatrix} = 0.$$

Using Laplace's rule to expand the determinant in terms of the elements of the top row gives

$$(4-\lambda)[(5-\lambda)(-10-\lambda)+60] - 7[6(-10-\lambda)+48] \\ + 6[-60+8(5-\lambda)] = 0,$$

which becomes

$$\lambda^3 + \lambda^2 - 4\lambda - 4 = 0. \tag{29}$$

Given that one of the roots is 2, we factorise this to give

$$(\lambda-2)(\lambda^2+3\lambda+2) = 0$$

or

$$(\lambda-2)(\lambda+2)(\lambda+1) = 0,$$

hence the eigenvalues are -2, -1 and 2.

Solution to Exercise 27

(a) For the matrix in Exercise 26(a), $\operatorname{tr}\mathbf{A} = 1 + 7 - 6 = 2$. Comparing this with the sum of the eigenvalues $\lambda_1 + \lambda_2 + \lambda_3 = -1 + 1 + 2 = 2$, we see that they are equal.

We could calculate $\det\mathbf{A}$ by hand. But instead let us note that from the equations leading up to equation (28), we have

$$\det(\mathbf{A} - \lambda\mathbf{I}) = -(\lambda^3 - 2\lambda^2 - \lambda + 2).$$

Setting $\lambda = 0$ then gives $\det\mathbf{A} = -2$. Comparing this with the product of the eigenvalues $\lambda_1 \times \lambda_2 \times \lambda_3 = -1 \times 1 \times 2 = -2$, we see that they are equal.

(b) For the matrix in Exercise 26(b), comparing $\operatorname{tr}\mathbf{A} = 4 + 5 - 10 = -1$ with $\lambda_1 + \lambda_2 + \lambda_3 = -2 - 1 + 2 = -1$, we see that they are equal.

Also, from equation (29) and the equations leading up to it, $\det(\mathbf{A} - \lambda\mathbf{I}) = -(\lambda^3 + \lambda^2 - 4\lambda - 4)$, so setting $\lambda = 0$ we get $\det\mathbf{A} = 4$. Comparing this with $\lambda_1 \times \lambda_2 \times \lambda_3 = -2 \times -1 \times 2 = 4$, we see that they are equal.

Solution to Exercise 28

Since $\mathbf{A}$ is a real matrix, another eigenvalue must be the complex conjugate of λ_1, i.e. $\lambda_2 = \overline{\lambda_1} = 1 - i$.

Further, the trace rule $\lambda_1 + \lambda_2 + \lambda_3 = \operatorname{tr}\mathbf{A}$ gives

$$(1 + i) + (1 - i) + \lambda_3 = 3,$$

hence $\lambda_3 = 1$.

Solution to Exercise 29

Because the matrix is triangular, the eigenvalues are 1 and 2.

Solution to Exercise 30

For the eigenvalue to be repeated, we require $\sqrt{(a + d)^2 - 4(ad - b^2)} = 0$, i.e. $(a - d)^2 + 4b^2 = 0$. This is true only if $a = d$ and $b = 0$, so the only symmetric 2×2 matrices with a repeated eigenvalue are of the form $\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$.

Solution to Exercise 31

(a) The eigenvalues are real, since $\mathbf{A}$ is real and symmetric. One is positive and the other negative, since $\lambda_1\lambda_2 = \det\mathbf{A} < 0$. Also, $\lambda_1 + \lambda_2 = \operatorname{tr}\mathbf{A} = 50$.

(b) The eigenvalues are the diagonal entries 67 and -17, since $\mathbf{A}$ is triangular.

(c) The eigenvalues are real, since $\mathbf{A}$ is real and symmetric. In fact, $\mathbf{A}$ is non-invertible, since $\det\mathbf{A} = 0$. Thus one eigenvalue is 0. Hence the other is 306, since $0 + \lambda_2 = \operatorname{tr}\mathbf{A} = 306$.

Solution to Exercise 32

(a) The characteristic equation is

$$\begin{vmatrix} 8-\lambda & -5 \\ 10 & -7-\lambda \end{vmatrix} = 0.$$

Expanding this gives $(8-\lambda)(-7-\lambda)+50=0$, which simplifies to $\lambda^2-\lambda-6=0$. So the eigenvalues are $\lambda=3$ and $\lambda=-2$.

Let $\mathbf{v}=[x \quad y]^T$ be an eigenvector.

- For $\lambda=3$, the eigenvector equations (20) and (21) become

  $$5x-5y=0 \quad \text{and} \quad 10x-10y=0,$$

  which reduce to the single equation $y=x$. So (setting $x=1$) an eigenvector corresponding to $\lambda=3$ is $[1 \quad 1]^T$.

- For $\lambda=-2$, the eigenvector equations become

  $$10x-5y=0 \quad \text{and} \quad 10x-5y=0,$$

  which reduce to the single equation $y=2x$. So (setting $x=1$) an eigenvector corresponding to $\lambda=-2$ is $[1 \quad 2]^T$.

(b) The characteristic equation is

$$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = 0.$$

Expanding this gives $(2-\lambda)^2-1=0$, which simplifies to $\lambda^2-4\lambda+3=0$. So the eigenvalues are $\lambda=3$ and $\lambda=1$.

Let $\mathbf{v}=[x \quad y]^T$ be an eigenvector.

- For $\lambda=3$, the eigenvector equations (20) and (21) become

  $$-x+y=0 \quad \text{and} \quad x-y=0,$$

  which reduce to the single equation $y=x$. So (setting $x=1$) an eigenvector corresponding to $\lambda=3$ is $[1 \quad 1]^T$.

- For $\lambda=1$, the eigenvector equations become

  $$x+y=0 \quad \text{and} \quad x+y=0,$$

  which reduce to the single equation $y=-x$. So (setting $x=1$) an eigenvector corresponding to $\lambda=-1$ is $[1 \quad -1]^T$.

Solution to Exercise 33

The eigenvector equations are

$$\begin{aligned} (\cos\theta-\lambda)x - \qquad (\sin\theta)y &= 0, \\ (\sin\theta)x + (\cos\theta-\lambda)y &= 0. \end{aligned}$$

- For $\lambda=\cos\theta+i\sin\theta$, the eigenvector equations become

  $$-(i\sin\theta)x-(\sin\theta)y=0 \quad \text{and} \quad (\sin\theta)x-(i\sin\theta)y=0,$$

  which reduce to the single equation $iy=x$ (since $\sin\theta\neq 0$ as θ is not an integer multiple of π). So setting $x=i$, a corresponding eigenvector is $[i \quad 1]^T$.

(Had we set $x = 1$, the eigenvector would be $[1 \quad -i]^T$, which is equally valid as it is just the first eigenvector multiplied by $-i$.)

- For $\lambda = \cos\theta - i\sin\theta$, the eigenvector equations become

$$(i\sin\theta)x - (\sin\theta)y = 0 \quad \text{and} \quad (\sin\theta)x + (i\sin\theta)y = 0,$$

which reduce to the single equation $-iy = x$ (since $\sin\theta \neq 0$), so a corresponding eigenvector is $[-i \quad 1]^T$ or any scalar multiple.

Solution to Exercise 34

(a) The matrix is upper triangular, so the eigenvalues are 2, -3 and 4.

The eigenvector equations for $\mathbf{v} = [x_1 \quad x_2 \quad x_3]^T$ are

$$\begin{aligned} (2-\lambda)x_1 + \quad x_2 - \quad x_3 &= 0, \\ (-3-\lambda)x_2 + \quad 2x_3 &= 0, \\ (4-\lambda)x_3 &= 0. \end{aligned}$$

- For $\lambda = 2$, the eigenvector equations become

$$x_2 - x_3 = 0, \quad -5x_2 + 2x_3 = 0, \quad 2x_3 = 0,$$

which reduce to $x_2 = x_3 = 0$. If we assign $x_1 = k$, for arbitrary non-zero k, then a corresponding eigenvector is $k[1 \quad 0 \quad 0]^T$. Choosing $k = 1$ gives $\mathbf{v} = [1 \quad 0 \quad 0]^T$.

- For $\lambda = -3$, the eigenvector equations become

$$5x_1 + x_2 - x_3 = 0, \quad 2x_3 = 0, \quad 7x_3 = 0,$$

which reduce to $5x_1 + x_2 = 0$ and $x_3 = 0$. So assigning $x_1 = k$, for arbitrary non-zero k, we get a corresponding eigenvector $k[1 \quad -5 \quad 0]^T$. Choosing $k = 1$ gives $\mathbf{v} = [1 \quad -5 \quad 0]^T$.

- For $\lambda = 4$, the eigenvector equations become

$$-2x_1 + x_2 - x_3 = 0, \quad -7x_2 + 2x_3 = 0, \quad 0 = 0.$$

Choosing $x_3 = 14$ keeps the numbers simple, and a corresponding eigenvector is $\mathbf{v} = [-5 \quad 4 \quad 14]^T$.

(b) The characteristic equation is

$$\begin{vmatrix} -\lambda & 2 & 0 \\ -2 & -\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = 0.$$

To simplify the evaluation of the determinant, we interchange the first and third rows (remember that this just changes the sign of the determinant). This gives the characteristic equation as $(1-\lambda)(\lambda^2 + 4) = 0$, so the eigenvalues are 1, $2i$ and $-2i$.

The eigenvector equations for $\mathbf{v} = [x_1 \quad x_2 \quad x_3]^T$ are

$$\begin{aligned} -\lambda x_1 + 2x_2 \quad &= 0, \\ -2x_1 - \lambda x_2 \quad &= 0, \\ (1-\lambda)x_3 &= 0. \end{aligned}$$

- For $\lambda = 1$, the eigenvector equations become

$$-x_1 + 2x_2 = 0, \quad -2x_1 - x_2 = 0, \quad 0 = 0.$$

These give $x_1 = x_2 = 0$. So choosing $x_3 = 1$, a corresponding eigenvector is $[0 \quad 0 \quad 1]^T$.

- For $\lambda = 2i$, the eigenvector equations become

$$-2ix_1 + 2x_2 = 0, \quad -2x_1 - 2ix_2 = 0, \quad (1 - 2i)x_3 = 0,$$

which reduce to $x_2 = ix_1$ and $x_3 = 0$. So choosing $x_1 = 1$, a corresponding eigenvector is $[1 \quad i \quad 0]^T$.

- Similarly, an eigenvector corresponding to $\lambda = -2i$ is $[1 \quad -i \quad 0]^T$.

Solution to Exercise 35

The eigenvector equations for $\mathbf{v} = [x_1 \quad x_2 \quad x_3]^T$ are

$$\begin{aligned}(4-\lambda)x_1 + 7x_2 + 6x_3 &= 0,\\ 6x_1 + (5-\lambda)x_2 + 6x_3 &= 0,\\ -8x_1 - 10x_2 + (-10-\lambda)x_3 &= 0.\end{aligned}$$

- For $\lambda = -2$, the augmented matrix is

$$\left[\begin{array}{ccc|c} 6 & 7 & 6 & 0\\ 6 & 7 & 6 & 0\\ -8 & -10 & -8 & 0\end{array}\right].$$

In this case it is helpful to interchange rows before doing anything else, so the starting arrangement will be

$$\begin{array}{c}\left[\begin{array}{ccc|c} -8 & -10 & -8 & 0\\ 6 & 7 & 6 & 0\\ 6 & 7 & 6 & 0\end{array}\right]\end{array}\begin{array}{l}\mathbf{R}_1\\ \mathbf{R}_2\\ \mathbf{R}_3\end{array}.$$

Reducing the elements below the leading diagonal in column 1 to zero (and shortcutting the usual procedure by subtracting the two identical rows at this early stage):

$$\begin{array}{l} \\ \mathbf{R}_2 + \frac{3}{4}\mathbf{R}_1\\ \mathbf{R}_3 - \mathbf{R}_2\end{array}\left[\begin{array}{ccc|c} -8 & -10 & -8 & 0\\ 0 & -\frac{1}{2} & 0 & 0\\ 0 & 0 & 0 & 0\end{array}\right]\begin{array}{l}\mathbf{R}_1\\ \mathbf{R}_{2a}\\ \mathbf{R}_{3a}\end{array}.$$

The final row of zeros allows us to assign x_3 the arbitrary non-zero value k. The second row tells us that $x_2 = 0$, and back substituting gives $x_1 = -k$. Thus the general form of the eigenvector is $\mathbf{v} = k[1 \quad 0 \quad -1]^T$, where k is an arbitrary non-zero value.

- For $\lambda = -1$, the augmented matrix is

$$\left[\begin{array}{ccc|c} 5 & 7 & 6 & 0\\ 6 & 6 & 6 & 0\\ -8 & -10 & -9 & 0\end{array}\right]\begin{array}{l}\mathbf{R}_1\\ \mathbf{R}_2\\ \mathbf{R}_3\end{array}.$$

Reducing the elements below the leading diagonal in column 1 to zero:

$$\begin{array}{r} \\ 5\mathbf{R}_2 - 6\mathbf{R}_1 \\ 5\mathbf{R}_3 + 8\mathbf{R}_1 \end{array} \left[\begin{array}{rrr|r} 5 & 7 & 6 & 0 \\ 0 & -12 & -6 & 0 \\ 0 & 6 & 3 & 0 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array}.$$

Reducing the element below the leading diagonal in column 2 to zero:

$$\begin{array}{r} \\ \\ 2\mathbf{R}_{3a} + \mathbf{R}_{2a} \end{array} \left[\begin{array}{rrr|r} 5 & 7 & 6 & 0 \\ 0 & -12 & -6 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The final row of zeros allows us to assign x_3 the arbitrary non-zero value k. Back substitution then tells us that $x_2 = -\frac{1}{2}k$, and back substituting again gives $x_1 = -\frac{1}{2}k$. Thus the general form of the eigenvector is $\mathbf{v} = k[-\frac{1}{2} \quad -\frac{1}{2} \quad 1]^T$, where k is an arbitrary non-zero value. Choosing $k = 2$ gives $\mathbf{v} = [-1 \quad -1 \quad 2]^T$.

- For $\lambda = 2$, the augmented matrix is

$$\left[\begin{array}{rrr|r} 2 & 7 & 6 & 0 \\ 6 & 3 & 6 & 0 \\ -8 & -10 & -12 & 0 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \end{array}.$$

Reducing the elements below the leading diagonal in column 1 to zero:

$$\begin{array}{r} \\ \mathbf{R}_2 - 3\mathbf{R}_1 \\ \mathbf{R}_3 + 4\mathbf{R}_1 \end{array} \left[\begin{array}{rrr|r} 2 & 7 & 6 & 0 \\ 0 & -18 & -12 & 0 \\ 0 & 18 & 12 & 0 \end{array}\right] \begin{array}{l} \mathbf{R}_1 \\ \mathbf{R}_{2a} \\ \mathbf{R}_{3a} \end{array}.$$

Reducing the element below the leading diagonal in column 2 to zero:

$$\begin{array}{r} \\ \\ \mathbf{R}_{3a} + \mathbf{R}_{2a} \end{array} \left[\begin{array}{rrr|r} 2 & 7 & 6 & 0 \\ 0 & -18 & -12 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The final row of zeros allows us to assign x_3 the arbitrary non-zero value k. Back substitution then tells us that $x_2 = -\frac{2}{3}k$, and back substituting again gives $x_1 = -\frac{2}{3}k$. Thus the general form of the eigenvector is $\mathbf{v} = k[-\frac{2}{3} \quad -\frac{2}{3} \quad 1]^T$, where k is an arbitrary non-zero value. Choosing $k = 3$ gives $\mathbf{v} = [-2 \quad -2 \quad 3]^T$.

Solution to Exercise 36

(a) $\mathbf{s}_1^T\mathbf{s}_2 = (2)(3) + (1)(-6) = 0$, so the column vectors are orthogonal.

(b) $\mathbf{t}_1^T\mathbf{t}_2 = (2)(-2) + (2)(-2) + (1)(0) = -8$, so the column vectors are not orthogonal.

Solution to Exercise 37

The characteristic equation is

$$\begin{vmatrix} 5-\lambda & 2 \\ 2 & 2-\lambda \end{vmatrix} = 0.$$

Expanding the determinant gives $(5-\lambda)(2-\lambda) - 4 = 0$, hence the characteristic equation may be written as $\lambda^2 - 7\lambda + 6 = 0$, the roots of which are the eigenvalues $\lambda_1 = 6$ and $\lambda_2 = 1$. Clearly both eigenvalues are real, as they should be for a real symmetric matrix.

The eigenvector equation is

$$\begin{bmatrix} 5-\lambda & 2 \\ 2 & 2-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

- $\lambda_1 = 6$ gives the pair of equations

  $$-x + 2y = 0 \quad \text{and} \quad 2x - 4y = 0,$$

  which are equivalent to $x = 2y$. Choosing $y = 1$, we get $\mathbf{v}_1 = [2 \quad 1]^T$.
- $\lambda_2 = 1$ gives the pair of equations

  $$4x + 2y = 0 \quad \text{and} \quad 2x + y = 0,$$

  which are equivalent to $y = -2x$. Choosing $x = 1$, we get $\mathbf{v}_2 = [1 \quad -2]^T$.

The inner product is $\mathbf{v}_1^T\mathbf{v}_2 = 2(1) + (-2)1 = 0$, so the eigenvectors are orthogonal.

Solution to Exercise 38

Since $\mathbf{v}_1^T\mathbf{v}_1 = 4 + 1 = 5$ and $\mathbf{v}_2^T\mathbf{v}_2 = 1 + 4 = 5$, we have

$$\widehat{\mathbf{v}}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \widehat{\mathbf{v}}_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Solution to Exercise 39

From equation (24), $\mathbf{v} = (\widehat{\mathbf{v}}_1^T\mathbf{v})\widehat{\mathbf{v}}_1 + (\widehat{\mathbf{v}}_2^T\mathbf{v})\widehat{\mathbf{v}}_2$. Calculating the coefficients, we have

$$\widehat{\mathbf{v}}_1^T\mathbf{v} = \frac{1}{\sqrt{5}}[2 \quad 1]\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{3}{\sqrt{5}} \quad \text{and} \quad \widehat{\mathbf{v}}_2^T\mathbf{v} = \frac{1}{\sqrt{5}}[1 \quad -2]\begin{bmatrix} 1 \\ 1 \end{bmatrix} = -\frac{1}{\sqrt{5}}.$$

Hence

$$\mathbf{v} = \frac{3}{\sqrt{5}}\widehat{\mathbf{v}}_1 - \frac{1}{\sqrt{5}}\widehat{\mathbf{v}}_2.$$

We can easily check this by substituting the values for $\mathbf{v}_1$ and $\mathbf{v}_2$:

$$\mathbf{v} = \frac{3}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{1}{5}\begin{bmatrix} 1 \\ -2 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 5 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Acknowledgements

Grateful acknowledgement is made to the following sources:

Figure 16: This file is licensed under the Creative Commons Attribution-Share Alike Licence http://creativecommons.org/licenses/by-sa/3.0.

Figure 17: Larry Page / This file is licensed under the Creative Commons Attribution-Share Alike Licence http://creativecommons.org/licenses/by-sa/3.0.

Every effort has been made to contact copyright holders. If any have been inadvertently overlooked, the publishers will be pleased to make the necessary arrangements at the first opportunity.